Ontology and the Foundations of Mathematics

by

Gabriel Uzquiano

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1999

Author ....
Department of Linguistics and Philosophy
July, 29 1999

Certified by ....
Vann McGee
Professor of Philosophy
Thesis Supervisor

Accepted by ....
Vann McGee
Chairman, Department Committee on Graduate Students

Ontology and the Foundations of Mathematics

by

Gabriel Uzquiano

Submitted to the Department of Linguistics and Philosophy
on July, 29 1999, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Philosophy

Abstract

"Ontology and the Foundations of Mathematics" consists of three papers concerned with ontological issues in the foundations of mathematics. Chapter 1, "Numbers and Persons," confronts the problem of the inscrutability of numerical reference and argues that, even if inscrutable, the reference of the numerals, as we ordinarily use them, is determined much more precisely than up to isomorphism. We argue that the truth conditions of a variety of numerical modal and counterfactual sentences (whose acceptance plays a crucial role in applications) place serious constraints on the sorts of items to which numerals, as we ordinarily use them, can be taken to refer: Numerals cannot be taken to refer to objects that exist contingently such as people, mountains, or rivers, but rather must be taken to refer to objects that exist necessarily such as abstracta.

Chapter 2, "Modern Set Theory and Replacement," takes up a challenge to explain the reasons one should accept the axiom of replacement of Zermelo-Fraenkel set theory, when its applications within ordinary mathematics and the rest of science are often described as rare and recondite. We argue that this is not a question one should be interested in; replacement is required to ensure that the element-set relation is well-founded as well as to ensure that the cumulation of sets described by set theory reaches and proceeds beyond the level ω of the cumulative hierarchy. A more interesting question is whether we should accept instances of replacement on uncountable sets, for these are indeed rarely used outside higher set theory. We argue that the best case for (uncountable) replacement comes not from direct, intuitive considerations, but from the role replacement plays in the formulation of transfinite recursion and the theory of ordinals, and from the fact that it permits us to express and assert the (first-order) content of the modern cumulative view of the set-theoretic universe as arrayed in a cumulative hierarchy of levels.

Chapter 3, "A No-Class Theory of Classes," makes use of the apparatus of plural quantification to construe talk of classes as plural talk about sets, and thus provide an interpretation of both one- and two-sorted versions of first-order Morse-Kelley set theory, an impredicative theory of classes. We argue that the plural interpretation of impredicative theories of classes has a number of advantages over more traditional interpretations of the language of classes as involving singular reference to gigantic set-like entities, only too encompassing to be sets, the most important of these being perhaps that it makes the machinery of classes available for the formalization of much recent and very interesting work in set theory without threatening the universality of the theory as the most comprehensive theory of collections, when these are understood as objects.

Thesis Supervisor: Vann McGee
Title: Professor of Philosophy

Acknowledgments

My first debt is to George Boolos and to the three members of my committee, Vann McGee, Robert Stalnaker, and Michael Glanzberg. To those who know his work, it will be plain that I owe an enormous debt to George Boolos, whose writings have motivated a large part of my work in this dissertation. I had the incredible fortune and privilege of being his student, and I will always treasure the years I worked with him. Robert Stalnaker and Michael Glanzberg have provided detailed comments and objections to several drafts of this dissertation, and their criticism and support has made the end result far better than it would otherwise have been. I owe an immeasurable debt to Vann McGee, who directed this dissertation and whose careful and detailed commentaries on several parts of it have proved an invaluable source of insight; without his help, support, and encouragement, this dissertation could not have been written.

Two other members of the faculty, Richard Cartwright and Judith Thomson, have played a very important role in my work. The proposal defended in the last chapter of this dissertation was largely motivated by Richard Cartwright's writings and shaped by conversations with him. Judith Thomson has given me a tremendous amount of help with the first chapter of this dissertation, and her insights have proved invaluable. I should like to thank them both for their generosity with their time and ideas.

I should like to thank many other people who have commented on and criticized different parts of this dissertation: Stephen Yablo, Richard Heck, Alex Byrne, Ignasi Jané, Lisa Sereno, Matti Eklund, Miguel Hernando, Ólafur Páll Jónsson, Adam Elga, Elizabeth Harman, Patrick Hawley, Leonard Clapp, Ana Carolina Sartorio, Juan Comesaña, Roger White, and Ulrich Meyer. I am especially indebted to Agustín Rayo, whose enthusiasm about plural quantification has made my work on plural quantification and classes both better and more enjoyable.

Audiences at the sixth annual Harvard-MIT philosophy conference, and at the Philosophy departments of the universities of Yale and Rochester provided me with the opportunity to present several parts of this dissertation and provided a number of very helpful comments and valuable suggestions to refine them.

Much of the logic I know, I learned in the department of Logic, History and Philosophy of Science of the University of Barcelona. I should like to thank my professors and fellow students there, especially to Manuel García-Carpintero, Ignasi Jané, Josep Macià, and Manuel Pérez-Otero.

"La Caixa" provided me with financial support for my first two years at M.I.T, and I have received additional support from the Linguistics and Philosophy department. I am grateful for this financial support.

I would like to thank my family for their support over the years, and all of my friends who have made Cambridge and incredibly exciting place to live in. I cannot hope to enumerate them all, but, as one of them has put it, they probably know who they are.

Finally, what is perhaps my greatest debt is to Merche Fages who has shaped much of what I am today, and has been a constant source of love, support, and encouragement over the years. This dissertation is dedicated to her.

Contents

Chapter 1

Numbers and Persons

1.1 Introduction: The Problem

As reference is ordinarily conceived, in order for a term to refer, there must be some combination of intentions, central features of usage, and relevant non-linguistic facts that connect specific uses of the term with its referent. Provided that numerals refer, then, there must be some combination of intentions and central features of use of the numerals either in theory or in applications, which, in conjunction with relevant non-linguistic facts, such as the facts of arithmetic, connect specific uses of the numerals with their referents, the numbers.

A very simple permutation argument establishes, it seems to me, that, provided again that numerals refer, this is hopeless as a picture of numerical reference. If π is a permutation of the natural numbers, then a reinterpretation of the arithmetical vocabulary that takes the usual numeral for each number n to refer not to n itself but to $\pi(n)$ and makes compensatory changes in the interpretation of each arithmetical predicate will be indistinguishable from the original interpretation with respect to the truth values they assign to each numerical sentence. To illustrate this point, notice that one could exchange the referents of the numerals '4' and '5', provided only that obvious compensatory changes are made in the interpretation of each arithmetical predicate. For example, a predicate such as "is a successor of" would be reinterpreted to refer to a relation which is like the successor relation except for the fact that 5 bears that relation to 4 but not the reverse, and a functional expression such as "the cardinal number of" would be reinterpreted to refer to a function that assigns the number 5 to sets which contain exactly 4 members and the number 4 to sets which contain exactly 5 members.

Perhaps it will be replied that it is our intention to use the numeral '4' to refer to the number 4, and not to the number 5, that excludes the possibility of a perverse reinterpretation of the numerals of the sort we have just envisaged. But all the reply does is to postpone the problem; in order for us to intend the number 4, and not the number 5, to be the referent of a numeral, there must be some combination of mental factors and other relevant facts, such as the facts of arithmetic, that permit us to discern one from the other. But, unless mental states are individuated in terms of their objects, a minor variation on the permutation argument just given would seem to suggest that the mental states of someone with an intention to refer to the number 4 would be qualitatively indistinguishable in all the relevant respects from the mental states of someone with an intention to refer to the number onto which the number 4 is permuted, the number 5.

I think we should admit that the moral of the permutation argument is that, provided that numerals refer, there is no combination of intentions, central features of usage, and non-linguistic facts that may be used to discern one interpretation of the numerals from a variety of alternative interpretations that differ from it in the natural numbers they assign to some numerals. There are reasons to be concerned. A paradox immediately confronts us: For, even after we realize that the reference of the numeral '4' is inscrutable, what could be more evident that, provided that the number 4 exists, the numeral '4' refers in the language we actually speak to the number 4 and to nothing else? A solution to the paradox might be to treat reference disquotationally, regarding sentences such as "If 4 exists, then '4' refers in the language we actually speak to 4 and to nothing else" as analytic. Such a solution would perhaps help us to salvage the intuition that, provided that the number 4 exists, the numeral '4' refers in the language we actually speak to the number 4 and to nothing else, but it would certainly sever the tie between reference and the activities of speakers.[1] Be that as it may, it is not my intention in this chapter to confront this paradox, but rather to discuss another question that emerges with the inscrutability of numerical reference.

[1]The disquotational line of response to the inscrutability of mathematical reference is briefly discussed by Vann McGee in (McGee, 1993), 103-109.

Another, perhaps more serious difficulty that emerges with the inscrutability of numerical reference is due to the fact that the standard Tarskian semantics explains the truth conditions of complete sentences in terms of reference and satisfaction. Now: if numerals do not refer, then it surely will not be possible to explain the truth of va sentence like "$4 < 5$" in terms of reference and satisfaction.

This need not be an insurmountable problem, but it requires the development of an alternative explanation of the truth conditions of numerical sentences in terms other than reference and satisfaction. To develop a perfectly general alternative to the Tarskian account would, no doubt, be a task of staggering difficulty, but, fortunately, the numerical case is considerably simpler than the general case. Thus, when attention is restricted to the numerical case, it is not difficult to develop an alternative, attractive account of truth that explains the truth conditions of arithmetical, and other numerical sentences, without appeal to determinate reference. Provided that some combination of intentions, central, non-negotiable features of usage, and relevant, non-linguistic facts selects a class of candidate reference relations for the entire class of numerals given by isomorphic copies of the natural number system, we can take the entire sequence of numerals to co-vary over all the candidate referents, and count a numerical sentence true just in case the sentence is true on all the candidate interpretations.[2]

What are the candidate referents for the numerals? To settle this question is to discern the limits of the inscrutability of numerical reference, and that is precisely the aim of this chapter. Perhaps there is no combination of intentions, features of use of the numerals either in theory or in applications, and other relevant facts that connect them uniquely with their referents, but the question remains whether thcy are able to at least exclude certain items as candidates to be the referents of the numerals. After all, for all the permutation argument implies, the scope of the inscrutability of numerical reference may well be restricted to the domain of natural numbers.

Or is it? Are items of an ostensibly different sort from numbers candidates to be the

[2]This is no trivial assumption, but it is nonetheless one that I will take for granted for present purposes; I will assume, in particular, that we manage to select a class of models, given by isomorphic copies of the natural number system, candidate reference relations for the entire class of numerals.

referents of the numerals? Are sets candidate referents for the numerals, as ordinarily used in theory and in applications? Is the unit set of the number 3 a candidate referent for the numeral '4'? Are symbols candidate referents for the numerals as used in theory and applications? Is the Arabic numeral '4' a candidate referent for the numeral '4'?

Or are ordinary objects candidate referents for the numerals? Is, to vary an example from Richard Cartwright,[3] Mo Vaughn, the man who used to play first base for the 1998 Boston Red Sox, a candidate referent for the numeral '4', as ordinarily used in theory and in applications? Presumably not, for, if the numeral '4' referred to Mo Vaughn, then, by semantic descent, we could infer that the number 4 is identical with Mo Vaughn, and that is a rather extravagant claim to make.

There is a deservedly famous argument due to Paul Benacerraf that is often taken to broaden the scope of the inscrutability to a wide range of items other than natural numbers.[4] Benacerraf called attention to a problem that emerges in the context of ontological reduction. It is ordinarily assumed that mathematics, and, in particular, number theory reduces to set theory. This reduction proceeds via the identification of the natural numbers with specific sets, but, as Benacerraf reminded us, several methods can be used to that end. What are perhaps the two most common methods for reducing number theory to set theory are due to Ernst Zermelo, who identified each natural number with the singleton of its predecessor, and to John von Neumann, who identified each natural number with the set of its predecessors. Now: Benacerraf's crucial observation is that both plans seem to serve the purpose of the reduction equally well: no amount of reflection on our central uses of the arithmetical vocabulary either in theory or in applications or on the facts of arithmetic would seem to settle the question of whether numerals refer to finite Zermelo ordinals rather than with to Neumann's finite ordinals.

Benacerraf observed further that, in general, no amount of reflection on central uses of the arithmetical vocabulary in theory and in applications or on the facts of arithmetic appears to settle the question whether numerals refer to the members of one rather than

[3]In the Addenda to "Propositions" in (Cartwright, 1987, 52).

[4]Cf. (Benacerraf, 1965).

another isomorphic copy of the natural number system, regardless of their sort.[5] The conclusion suggests itself that the reference of the numerals is determined no more precisely than up to isomorphism.

This conclusion — that the reference of the numerals is determined no more precisely than up to isomorphism — may have been one Benacerraf's line of argument warranted, but it was certainly not the conclusion he extracted. Benacerraf never presented his argument as an argument for the inscrutability of numerical reference. Nor did he even subscribe to the thesis that the reference of the numerals is determined no more precisely than up to isomorphism; in 1965, he insisted that only isomorphic copies of the natural number system in which the relation "is less than" is recursive are candidates to play the role of the natural number system.[6] Instead, he concluded that, if numerals do not refer, then "if the truth be known, there are no such things as numbers; which is not to say that there are not at least two prime numbers between 15 and 17."

Whatever Benacerraf's conclusion may have been, a variety of authors have extracted the conclusion that numerical reference is determined no more precisely than up to isomorphism, and the interest of this thesis is that, when combined with the alternative account of truth conditions outlined before, it provides us with a quite attractive explanation of the truth conditions of numerical sentences: a sentence of arithmetic, for example, is true if and only if it is true in every isomorphic copy of the natural number system, or, equivalently, if and only if it is true in some isomorphic copy of the natural number system.

Of course it is an immediate consequence of the thesis that the reference of the numerals is determined no more precisely than up to isomorphism that Mo Vaughn is a perfectly suitable candidate to be the interpretation of the numeral '4'. For there are, after all, isomorphic copies of the natural number system in which the numeral '4' refers to Mo Vaughn. This fact, in combination with the account of truth conditions just now given, would seem to commit us, for example, with the rather extravagant claim that a sentence such as "The number 4 has recently moved to California" is not false.

[5] An isomorphic copy of the natural number system is an infinite sequence of objects each one of which has only finitely many predecessors.

[6] He has recently recanted in (Benacerraf, 1996).

What I would like to do in this chapter is to argue that the conclusion that Mo Vaughn is a candidate to be the referent of the numeral '4' is not only extravagant, it is false, too. There are central features of use of the numerical terms in applications that constrain the sorts of items to which we may take the numerals to refer; these features exclude, in particular, ordinary objects like people, mountains, or rivers as candidates to be the referents of the numerals. Therefore, though hopelessly inscrutable, the reference of the numerical terms is, I want to argue, certainly determined much more precisely than up to isomorphism.

There are at least two important reasons this thesis is of interest.

The first reason is peculiar to the philosophy of mathematics. One way to avoid the unsavory consequences of Platonism is to adopt a stance of Aristotelian realism, according to which mathematical objects are ordinary objects considered from an abstract point of view.[7] For Aristotle, geometrical objects were ordinary objects examined from an abstract point of view that ignored all aspects of them other than their size, shape and position. As attractive as the Aristotelian view may seem for elementary geometry, it must be admitted that it is not a tenable view to entertain with respect to contemporary mathematics. For one reason, it would seem thoroughly unreasonable to expect us to encounter in physical experience the bewildering variety of structures studied by contemporary set theory. This, however, is no reason to reject the Aristotelian stance with respect to arithmetic, according to which numbers are best viewed as ordinary objects considered from an abstract point of view, a point of view that ignores all aspects of them other than their position in a certain structure, the natural number system. But if there are limits on the inscrutability of numerical reference and ordinary objects turn out not to be candidates to be the referents of the numerals and other numerical terms, then the Aristotelian view with respect to arithmetic will no longer be tenable.

The second reason is much more general and concerns the theory of reference. There are a variety of arguments in the literature designed to establish that the inscrutability of reference is a perfectly general phenomenon, and not one restricted to the numerical case.

[7]Cf. Metaphysics, M.

The issue remains highly controversial, but Hilary Putnam has advanced a permutation argument, which, if correct, would even seem to establish that, whatever the subject matter, reference is never determined more precisely than up to isomorphism.[8] The point of Putnam's permutation argument is to establish that the truth conditions of complete sentences determine the reference of the terms and predicates of the language no more precisely than up to isomorphism. Since, for some philosophers, truth conditions are all there is to determine reference, they conclude that, in the end, reference is never determined more precisely than up to isomorphism. The thesis that, whatever the subject matter, reference is determined no more precisely than up to isomorphism remains highly controversial. For it is at least arguable that, in the case of ordinary reference, there are factors other than usage or truth conditions of complete sentences that constrain the reference of most of the terms we use; there are for example demonstrative identifications and causal connections that may constrain the reference of our terms and predicates.

The interest of the numerical case is that it is arguable that there are no factors — causal, demonstrative, or otherwise — other than truth conditions of complete sentences that can be cited to mitigate the extent of the inscrutability, and it would seem to present us with an unadorned case of a subject matter for which reference is determined no more precisely than up to isomorphism.

What I would like to suggest is that reflection on the considerably simpler case of numerical reference may throw light upon the general case, too. For, after all, if — and this is what we shall try to establish — numerical reference is determined more precisely than up to isomorphism, this will certainly be not due to the existence of demonstrative identifications or other causal factors which happen to help connect the numerals uniquely with their referents. Thus, it is at least conceivable that whatever constrains the reach of the inscrutability of numerical reference may well place constraints on the scope of ordinary reference, if there is such a perfectly general phenomenon, too.

[8]Cf. (Putnam, 1981). Davidson makes use of a similar permutation argument in (Davidson, 1979). Another example is Quine's argument from proxy functions. To the best of my knowledge, the first use of a permutation argument to draw a perfectly general point about reference is due to Richard Jeffrey who exchanged people with their social security numbers in (Jeffrey, 1964). I am grateful to Vann McGee for telling me of this reference.

1.2 The scope of the inscrutability of numerical reference

In the end, the suggestion I will want to make is that there are features of our use of the numerals in applications in modal and counterfactual contexts that exclude interpretations that assign to the numerals objects that exist contingently, such as Mo Vaughn.

To see this clearly, however, we need to review first what is nowadays a familiar distinction. Compare the two different uses of the numeral '4' in the sentences:

(1) Jupiter has exactly 4 major moons,

and

(2) T' number of Jupiter's major moons is 4.[9]

In (1), the numeral '4' acts as an adjective, not as a substantive, and, given the usual recursive definition of the numerical quantifiers,[10] (1) is often paraphrased into the language of first-order logic with identity as:

(1.a) $\exists_4 x$ (x is a major moon of Jupiter),

which can be unpacked as:

(1.b) $\exists x_1 \exists x_2 \exists x_3 \exists x_4 (\bigwedge_{1 \leq i < j \leq 4} x_i \neq x_j \wedge \forall y(y \text{ is a major moon of Jupiter} \leftrightarrow \bigvee_{1 \leq i \leq 4} y = x_i))$

In contrast with its occurrence in (1), the occurrence of the numeral '4' in (2) is referential; it refers to an object, the number 4. The truth value of (2) depends both on the existence of a particular number, the number 4, and on whether 4 is indeed the number of Jupiter's major moons.

[9]This is one of Frege's examples in (Frege, 1884), section 55. Frege held that numbers are objects and that "statements of number" like (1) had to be analyzed in terms of (2).

[10]The definition of the numerically definite quantifiers 'There is exactly 1', 'There is exactly 2', and the like is:
$\exists_0 x\ Fx \leftrightarrow \forall x \neg Fx$,
$\exists_{n+1} x\ Fx \leftrightarrow \exists y(Fy \wedge \exists_n x(Fx \wedge x \neq y))$.

In a recent paper, Neil Tennant, (Tennant, 1997), calls attention to the distinction just now drawn with a view to establishing the necessary existence of numbers. Tennant contends that the connection between (1) and (2) is that of analytic equivalence. Then of course, the analyticity of certain instances of:

(3) The number of F's $= 0 \leftrightarrow \forall x\ \neg Fx$

(4) The number of F's $= 1 \leftrightarrow \exists x \forall y\ (Fy \leftrightarrow x = y)$

(5) The number of F's $= 2 \leftrightarrow \exists x_1 \exists x_2\ (x_1 \neq x_2 \wedge \forall y\ (Fy \leftrightarrow x_1 = y \vee x_2 = y))$

would seem to provide us with straightforward enough arguments for the (necessary) existence of the numbers 0, 1, and 2. To establish the existence of the number 0, one needs only note that $\forall x\ \neg(x \neq x)$. For this, coupled with the relevant instance of (3), implies that the number of things that are not self-identical is 0. In like manner, one obtains the existence of the number [illegible]en F is instantiated with the predicate $(x = 0)$ in (4), and the existence of the num[illegible], when F is instantiated with $(x = 0 \vee x = 1)$ in (5). It should now be straightforward how to extend this line of argument to establish the existence of all the rest of natural numbers.[11]

It might appear that the line of argument just now outlined supplies us with an argument for the thesis that, as they are ordinarily used in applications, numerals cannot be taken to refer to persons and other contingent existents. If numbers are necessary existents, then, by semantic ascent, numerals cannot be taken to refer to contingent existents, which is the conclusion we wanted. Unfortunately, there are reasons to be dissatisfied.

Two difficulties deserve special mention.

First, it is doubtful that all biconditionals of the form of (3), (4), and (5) are analytic. The first point to be noticed is that, unless suitable restrictions are placed on the sorts of predicates with which F can be instantiated, it will be possible to produce biconditionals of the form of (5), for example, which are not only not analytic, but plainly false. For example,

---

[11] If the argument appears familiar, it is. The argument is modeled after Frege's proof in the *Foundations* that the number 0 exists and that every number has a successor. I should notice, however, that, as Tennant outlines his argument in (Tennant, 1997), it relies only on (3) and on two other principles concerning the existence of successors.

to employ a Fregean example, suppose we instantiate F in the relevant biconditional with the predicate 'drew the king's carriage':

(6) The number of horses that drew the king's carriage $= 4 \leftrightarrow$
$\exists x_1 \exists x_2 \exists x_3 \exists x_4 (\bigwedge_{1 \leq i < j \leq 4} x_i \neq x_j \wedge \forall y (y$ is a horse that drew the king's carriage $\leftrightarrow$
$\bigvee_{1 \leq i \leq 4} y = x_i))$

On the face of it, (6) is ambiguous between a distributive and a perhaps more salient, collective reading of the predicate "drew the king's carriage". Now, contrary to what Frege himself may have intimated in the *Foundations*, it may be true, for example, that the number of horses that drew the king's carriage is 4, but nevertheless true of no horse that it drew the king's carriage.[12] If so, then, on a collective interpretation of the predicate "drew the king's carriage," the conditional from left-to-right of (6) is not just not analytic, it is outright false.

But even if suitable restrictions are placed on the sorts of predicates with which F can be instantiated, one may still seriously doubt, for example, that the conditional from right-to-left of (3), which states that if there are no F's then there is such an object as the number 0, is analytic. And the reason is that, like many others, one may subscribe to the traditional picture of analytic truths as truths that lack content and make no substantive claims or commitments about the way the world is; and, in particular, as truths that don't entail the existence of particular objects. Indeed, not only do the relevant conditionals entail the existence of objects, by Frege's argument in the *Foundations*, they deliver the existence of an infinite number of objects as a consequence, which an added source of concern for someone who subscribes to the traditional picture of analyticity.

It will perhaps be replied that the Frege-Tennant strategy can be sustained in the presence of a weaker assumption, such as, for example, the assumption that biconditionals of the form of (3), (4) and (5) are necessary truths (just not analytic.) Perhaps so, but I, for one, know of no direct argument for the necessity of those biconditionals. And thus deriving the necessary existence of numbers from the necessity of biconditionals of the form

[12](Cf. Frege, 1884, section 46). I am grateful to Richard Cartwright for calling this reference to my attention.

of (3), (4) and (5) would be no better than deriving the necessary existence of numbers directly from the necessity of the Peano postulates. For, after all, what warrant we have for confidence that the axioms of arithmetic are necessary truths? Presumably, if there is a possible world in which there are no numbers, then the axioms of arithmetic are not necessary truths.

The second reason to be dissatisfied with Tennant's line of argument concerns the scope of the conclusion it warrants. It seems to me that Tennant's argument establishes its conclusion on the assumption that numerals are to be viewed as rigid designators, as terms which cannot be taken to refer to different objects in different possible worlds. But notice that the possibility that '4' refers to different objects in different possible worlds is not completely outrageous. As Russell conceived of the number 4, it consisted of the class of all sets containing exactly 4 members. If, as it is commonly assumed, a set can exist only in those possible worlds in which their members exist, then, on the Frege-Russell account of number, the numeral '4' refers to different classes in different possible worlds. Whereas, in the actual world, the class of 4-membered sets contains sets containing Mo Vaughn as an element, in other possible worlds in which Mo Vaughn doesn't exist, the class of 4-membered sets doesn't contain any set containing Mo Vaughn as a member.

Therefore, unless it is assumed that numerals are indeed genuine names, and hence rigid designators, the argument given by Tennant doesn't by itself establish that numbers are objects that exist necessarily, and hence that they are different from persons and other ordinary physical objects.

No matter, I think that there is another perfectly general line of argument for the conclusion that numerals cannot be taken to refer to contingent existents. The argument proceeds from the observation that it is an important, non-negotiable feature of our use of the numerals in elementary cardinal applications that, in general, when we count, it is enough for us to be entitled to assertion that the (cardinal) number of F's is n that there be exactly n F's. For example, when we count Jupiter's major moons, we assert that the number of Jupiter's major moons is 4 merely on the grounds that there are exactly 4 major moons in Jupiter's orbit.

It is not just that we incline — for whatever accident of psychology — to make the inference from "There are exactly n F's" to "The number of F's is n," but rather that this pattern of inference plays a crucial role in certain contexts of application. One reason for this concerns the expressive gains substantival uses of the numerals — as exemplified in (2) — provides to us. For example, it is the use of the numerals as substantives that permit us to represent the fact that the ratio of F's to G's is 4 in finite compass. The adjectival use of the numerals as adjectives provides us, at most, with the resources necessary to express that fact as an infinite disjunction of sentences like: "There is exactly 1 F and 4 G's," "There are exactly 2 F's and 8 G's," and the like, but certainly not those necessary to represent the fact that the ratio of F's to G's is 4 in finite compass.

The other reason it is important to us to be able to pass from instances of "There are exactly n F's" to instances of "The number of F's is n" is that it is this pattern of inference that makes the facts of arithmetic relevant for deductive purposes. It is such an inference that allows us to make use of the fact that 4 times 4 is 16 to conclude that the ratio of F's to G's is 4 in the case in which there are exactly 4 F's and 16 G's.

A word of caution is in order. I think it is important to distinguish the observation that, in general, we make the inference from "There are exactly n F's" to "The number of F's is n" from the claim that biconditionals of the form of (3), (4), and (5) are necessary truths, which would allow us to use a Frege-Tennant style of argument to conclude the necessary existence of numbers. The observation just made concerns a central trait of our practices in the use of the numerals in elementary cardinal applications, but remains neutral as to whether, for example, all instances of (3), (4), and (5) are, in fact, true. Admittedly, we are all inclined to think they are true, but, by itself, our disposition to assent to them doesn't quite make them true.

Nor is the observation under discussion supposed to refute the viability of nominalist projects aimed to encode much of what we achieve with the help of substantival uses of the numerals in more metaphysically benign terms; it is not my business here to refute the possibility that such a project can ever succeed to accomplish its aim. To repeat, the aim of this chapter is not to argue that ordinary usage can, in the end, be sustained, or perhaps recast in more metaphysically benign terms, but rather to discern what it takes for our central practices in the use of the numerals to be sustained. And, in particular, what are the constraints these central uses of the numerals place on the sorts of items to which we may take the numerals to refer.

To continue, then, numerals are used substantivally not only to describe actual circumstances, but to describe counterfactual circumstances, too; just as the sentence "The number of Jupiter's moons is 4" is used to describe the actual circumstance in which Jupiter has exactly 4 major moons, we can use a similar sentence to describe a counterfactual circumstance in which Jupiter has exactly 3 major major moons. In particular, we can describe a counterfactual circumstance in which Jupiter has exactly 3 major moons as a circumstance in which the number of Jupiter's major moons would have been 3. And, in general, a counterfactual circumstance in which there are exactly n F's as a circumstance in which the number of F's would have been n.

It is important to note that this is again not a dispensable feature of our use of the numerals in applications; again, it is not just that we are inclined to conceive of two different ways of describing certain counterfactual circumstances as interchangeable. Quite often, it is precisely the ability to pass from instances of "There are n F's" to instances of "The number of F's is n" in counterfactual contexts that allows us to reason about what would be the case in some counterfactual circumstance or another.

Now: on the assumption that numerals are rigid designators in that they do not refer to different objects in different possible worlds, it is not difficult to realize that a large number of inferences from "There are exactly n F's" to "The number of F's is n" in counterfactual contexts could not be sustained if they referred to objects that existed contingently.

Let me explain this point carefully. Unlike Jupiter, the planet Uranus has exactly 5 major moons: Ariel, Miranda, Oberon, Titania, and Umbriel. Since we could certainly use the numeral '4' to describe a counterfactual circumstance in which there are exactly 4 major moons of Uranus as a circumstance in which the number of Uranus' major moons is 4, there are serious constraints on the sorts of objects to which we may take the numeral '4' to refer. For example, the numeral '4' cannot refer to Ariel, one of Uranus' major moons. The reason is that if Ariel had not existed, Uranus would have had exactly 4 major moons, and thus our practices in ascriptions of number in counterfactual circumstances sanctions the assertion:

(7) If Ariel had not existed, the number of Uranus' major moons would have been 4.

But, from (7), we can infer:

(8) The number 4 would have existed, even if Ariel had not existed.

Therefore,

(9) $\Diamond$ ($4 \neq$ Ariel),

which, coupled with the principle of necessity of identity:

(NI) $\forall x \forall y\ (x = y \rightarrow \Box(x = y))$,

modulo the assumption that the numeral '4' is rigid, delivers:

(10) $4 \neq$ Ariel.

which is of course the conclusion that we wanted.

Thus far this is no better than Tennant's argument, for it, too, rests upon the assumption that numerals cannot be interpreted to refer to different objects in different possible worlds. The difference, however, is that there is a variation on the argument from (7) to (10) that dispenses with the assumption that the numerals are rigid in that they cannot be taken to refer to different objects in different possible worlds. For notice that it is not just that the inference from "Jupiter has 4 major moons, and it would have still had 4 major moons even if Mo Vaughn had not existed" to: "The number of Jupiter's major moons is 4, and 4 would have been the number of Jupiter's moons even if Mo Vaughn had not existed" is important to us, in general, we regard ourselves as entitled to the claim that, whatever the number of Jupiter's moons, it would have numbered Jupiter's moons even if Mo Vaughn had not existed. But now, from the premise:

(11) There is a number that is the number of Jupiter's major moons and which would have been the number of Jupiter's moons even if Mo Vaughn had not existed,

we can infer:

(12) There is a number that is the number of Jupiter's major moons and which would have existed even if Mo Vaughn had not existed.

Therefore,

(13) There is a number that is the number of Jupiter's major moons and which might have been different from Mo Vaughn,

which, again, coupled with the necessity of identity, (NI), yields:

(14) There is a number that is the number of Jupiter's major moons and which is different from Mo Vaughn.

The crucial observation now is that the truth of both (14) and the sentence "The number of Jupiter's moons is 4" are jointly incompatible with the assumption that the numeral '4' refers to Mo Vaughn, which, once again, is the conclusion we wanted.

It should now be evident that parallel arguments establish that the numeral 4 doesn't refer to other contingent existents such as Jupiter's moon Europa or to the river Charles. Indeed, if no contingent existent is the referent of the numeral '4', then the numeral '4' must refer to a necessary existent. And there is nothing special about the numeral '4': it is not difficult to come up with analogous arguments for the conclusion that the numeral '3' doesn't refer to objects that exist contingently such as Mo Vaughn, Europa, or King Juan Carlos. And if we can show that no contingent existent is the referent of the numeral '3', then the numeral '3' must refer to a necessary existent, too. And it should now be obvious how to extend these considerations to apply to the rest of numerals.

There is, then, some reason to think that in order for an item to be a candidate to be the referent of a numeral, it must itself be a necessary existent. Some reason, but perhaps not every reason. For some philosophers will perhaps be disturbed by the very idea of a necessary existent, perhaps because they believe that all there is are ordinary objects such as persons, mountains and rivers, perhaps because they believe that necessary existents are mysterious objects, and that we had better forego them. Whatever the reason, it is plain that the claim that numbers are objects that exist necessarily is a pretty strong claim by the lights of those who believe that no object exists as a matter of metaphysical necessity, so strong, they may say, that it calls into question the arguments of this section.

These philosophers will of course be utterly unimpressed by arguments like that from (7) to (10) on the grounds that numerals are better viewed as non-rigid terms that may be taken to refer to different objects in different possible worlds. And they will similarly object to arguments like that from (11) to (14) by proclaiming that, despite appearances, sentence (11) is either plainly false or true, but not for the reasons we have assumed.

There are two main lines such a philosopher may take in order to resist arguments like that from (11) to (14).

One strategy is to proclaim that, despite appearances to the contrary, (11) is false. Doubtless, it will not be open to them to deny that we do assent to (11). Nor will they be able to deny that we invariably infer (11) from: whatever the number of Jupiter's moons is, it would have remained the number of Jupiter's moons even if Mo Vaughn hadn't existed. Nevertheless, they might still question the need to take the fact that we assent to (11) at face value, and propose instead to encode a sentence like (11) as:

(15) There is a number that is the number of Jupiter's major moons, and if Mo Vaughn had not existed, there would have been a number which would have been the same number as the actual number of Jupiter's major moons is and would have numbered Jupiter's major moons.

where the relation "the same number as" is of course not to be confused with the identity relation, but is rather construed as a transworld relation that holds between referents of the numerals in different possible worlds. Thus for example, if '4' refers to Mo Vaughn in the actual world, but to Nomar Garciaparra in another possible world, w, which is as much like the actual world as possible except for the fact that Mo Vaughn doesn't exist in w, then 4, i.e., Mo Vaughn, is the number of Jupiter's major moons and, in w, where Mo Vaughn doesn't exist, the same number, 4, i.e., Nomar Garciaparra, is the number of Jupiter's moons.

This strategy has the advantage that it would seem to encode much of ordinary usage in metaphysically benign terms that do not require numbers to be necessary existents. All I can do in the face of this is to repeat that we do make the inference from: "whatever the number of Jupiter's moons is, it would have remained the number of Jupiter's moons, even if Mo Vaughn had not existed" to (11), that we hold (11) true, and that to recast (11) in terms of (15) would be to distort an important feature of our use of the numerals and other number words in modal and counterfactual contexts. Therefore, if one takes our use of the numerals and other numerical terms at face value, then she has to conclude that they refer to necessary existents. Moreover, since I am inclined to think that there may well be objects that could not have failed to exist, and I know of no independent reasons to refuse to take our use of number words at face value, I stand by my claim that the arguments presented here are evidence for the non-identity of numbers with objects that exist contingently such as people, mountains, and rivers.

The other line of response to the argument from (11) to (16) I would like us to consider is to admit that (11) is true, but to deny that it is true for the reasons one would normally assume.[13] (11) is true, the reply continues, not because the number of Jupiter's moons exists in possible worlds in which Mo Vaughn doesn't exist, but rather because "is the number of" is (or can be) a transworld relation. Thus it may be claimed that (11) *might* be read as:

(16) The actual number of Jupiter's major moons is such that, in the actual world, it would have numbered Jupiter's major moons even in a circumstance in which Mo Vaughn had not existed.

More can be said: unless the number 4 is identical with one of Jupiter's major moons, it is true that, in the actual world, the number 4 would have numbered Jupiter's major moons even in a circumstance in which the number 4 itself had not existed.

This alternative interpretation may perhaps be motivated with the help of an analogy. For consider what (17) says:

(17) Jupiter is brighter than whatever planet would have been closest to Saturn had Jupiter not existed.

[13] Thanks are due here to Stephen Yablo.

It may be argued that the claim made by (17) is not of course the non-sensical claim that Jupiter, in a possible world as much like the actual as possible in which Jupiter doesn't exist, is brighter than whatever planet happens to be closest to Saturn. Rather, (17) may be taken to make a transworld comparison: Jupiter in the actual world is brighter than whatever planet is closest to Saturn in a possible world in which Jupiter doesn't exist.

Likewise, then, the claim made by (16) may be taken to involve a transworld relation between a number, which may or may not be Mo Vaughn, and Jupiter's major moons as they exist in a possible circumstance in which Mo Vaughn doesn't exist. The move may be resisted by denying, in general, that sentences like (16) and (17) can be read to involve a transworld relation. Or, more modestly, by denying that a sentence like (16) can be read in the way just now suggested. Though I am inclined to entertain doubts about the availability of such an interpretation, I think that it would be somewhat unsatisfying to rest a rejoinder on such reservations.

Fortunately, I think there is a different line of response. This alternative line proceeds from the observation that the line of argument of this section extends to arguments based on counterfactuals, which, first, don't involve the relation "is the same number as," and, second, cannot be read in the manner suggested by the objection. An example of common such counterfactuals can be drawn from public key cryptography, which nowadays makes network communications secure.[14] A public key is a number a message recipient sends to a sender to allow the latter to encrypt the message to be sent. This number is generally the product of two very large primes, and therefore it is extremely difficult to factor. Now: once the sender receives the public key, she uses some mathematical function to scramble the message, and proceeds to send the result to the message recipient. At this point, the message can only be decrypted by someone who knows the original primes; in this case, the message recipient. This is a relatively straightforward arithmetical application, but observe that we are indeed committed to the truth of a counterfactual such as:

(18) There are two primes which would decrypt the message, that is, would generate a certain value as arguments of a certain mathematical function, even if Jupiter's

[14] A similar example was suggested to me by Michael Glanzberg.

moon Europa ceased to exist on Thursday, July 29 at 4 pm.

This conterfactual provides us with a guarantee that the message would be decrypted by the two large primes in questions even in a circumstance in which Jupiter's moon Europa ceased to exist on Thursday, July 29 at 4 pm. Otherwise, public key cryptography would be useless for practical purposes; as a message recipient, I would need to check a variety of facts concerning which objects exist in which circumstances in order to make sure that I would still be able to decode the message in each circumstance.

The important point to be noticed is that, from (18), a few steps will lead us to an argument whose conclusion is incompatible with the assumption that Europa is the referent of one of the numerals for the two primes in question. Yet, (18) cannot be read to make a claim about a transworld relation that connects two very large primes in the actual world with the value of a certain mathematical function in some other possible world.

This is not to deny that (18) could be recast in the manner suggested by the line of objection we considered before as:

(19) There are two prime numbers that decrypt the message, and if Jupiter's moon Europa ceased to exist on Thursday, July 29 at 4 pm., there would be two numbers which would be the same numbers as the two prime numbers that decrypt the message and which would decrypt the message.

However, the comments made before in response to that general line apply to this particular case, too.

I think we should conclude that the arguments of this section provide us with some reason for the conclusion that numerals cannot be taken to refer to objects that exist contingently. As I have stated it, this conclusion excludes ordinary objects as the sort of items to which numerals may not be taken to refer, but it also extends to a variety of objects commonly labeled "abstract." For example, consider the singleton of Mo Vaughn. If one assumes that Mo Vaughn's singleton can only exist in possible worlds in which Mo Vaughn does, then the argument that results from replacing the occurrences of the term 'Mo Vaughn' by the term '{Mo Vaughn}' in (11), (12), (13), and (14) establishes the non-identity of the number 4 with Mo Vaughn's singleton. Similar considerations apply to other abstract objects that exist contingently, such as the Red Sox or the American League.

The arguments presented thus far were aimed to establish that, provided that numerals refer, they refer to necessary existents. Yet, there are a variety of features people often ascribe to numbers other than necessary existence; numbers are often labeled "abstract," claimed to lack specific spatio-temporal location, and some discussions even suggest that numbers, and other mathematical objects, have little by way of intrinsic structure. I would like to end this section with the observation that the analogy between temporality and modality suggests that the general line of argument presented here can be extended to exclude the possibility that numerical terms refer to objects that exist at a time, but not at another. The analogy, in particular, suggests a temporal counterpart of the argument from (11) to (14):

(11') There is a number that is the number of Jupiter's major moons and which was the number of Jupiter's moons even in 1944, when Mo Vaughn did not yet exist,

we can infer:

(12') There is a number that is the number of Jupiter's major moons and which existed even in 1944, when Mo Vaughn did not yet exist.

Therefore,

(13') There is a number that is the number of Jupiter's major moons and which, in 1944, was different from Mo Vaughn,

which, again, coupled with a principle that gives partial expression to what may be called the eternity of identity:

(EI) $\forall x \forall y\ (\mathcal{P}(x = y) \rightarrow x = y)$,

where $\mathcal{P}$ is the temporal operator "It has been the case that,"[15] delivers:

[15] The other principle that gives expression to the eternity of identity is: $\forall x \forall y\ (\mathcal{F}(x = y) \rightarrow x = y)$, where $\mathcal{F}$ is the operator: "It will be the case that."

(14') There is a number that is the number of Jupiter's major moons and which is different from Mo Vaughn.

The important point to be noticed now is that the truth of both (14') and the sentence "The number of Jupiter's moons is 4" are jointly incompatible with the assumption that the numeral '4' refers to Mo Vaughn, which, once again, is the conclusion we wanted.

It is an interesting question whether a similar line of argument could be used to that numbers are not the sorts of objects that exist at a place but not at another, but note that such an argument would be importantly different from other, more traditional arguments from the premise that numbers are necessary existents to the conclusion that they lack specific spatio-temporal location. Here is an example of a different argument to the effect that if an individual has a certain spatio-temporal location, then it is not a necessary existent: if an individual, x, has a certain spatio-temporal location in the actual world, then we can surely imagine a possible world, w, as much as possible like the actual world with the difference that, in w, nothing occupies the location of x in the actual world. It seems right to say that x doesn't exist in w, and therefore that there is a possible world in which x doesn't exist. Therefore, if an individual x is spatio-temporally located, then x is not a necessary existent. But then, since numbers are necessary existents, then, by the preceding argument, they must lack a spatio-temporal location. What is the force of this argument is, however, a question which I shall not address here; suffice it to say that suspiciously similar considerations are sometimes employed to cast doubt upon the assumption that there are objects that exist necessarily: it would seem that one can always imagine an object away; to conceive of a circumstance in which the object doesn't exist. But then, if that circumstance constitutes a genuine possibility, it would seem that it is a possible world in which the object under consideration does not exist.

1.3 Three remarks on the general line of argument

I have claimed that there are central features of our use of the numerals that place serious constraints on the sorts of objects to which we may take the numerals to refer. In particular, I have argued that it is not open to us to interpret the numerals to refer to objects that exist contingently, such as people, mountains, or rivers. For, otherwise, our practices in the use of the numerals to describe counterfactual circumstances could not be sustained; sentences (8) and (13) are cases in point. This, I claim, is what the former line of argument establishes. It is now time to comment on what it is not aimed to — and, I think, does not — establish.

Two remarks on what the former line of reasoning is not aimed to accomplish:

First, it might be thought that the connection between sentences of the form "There are n F's" and sentences of the form "The number of F's is n" provides us with a priori grounds for believing in the (necessary) existence of numbers. For example, it might be suggested that the connection between the sentence "There are no objects that are not self-identical" and the sentence "The number of objects that are not self-identical is 0" provides us with a priori grounds for believing that the number 0 exists. After all, we know a priori that there are no objects that are not self-identical, and we all seem disposed to assert that the number of objects that are not self-identical is 0 simply on the grounds that there are no such objects.

Should we conclude that we know a priori that the number 0 exists?

If we do, then we can make use of the line of argument outlined by Tennant to infer the (necessary) existence of the rest of numbers on a priori grounds, too.

Now, I want to stress that, whatever else it may achieve, the line of argument given here doesn't provide us with a priori grounds for believing in the (necessary) existence of numbers. For let us look more closely at the biconditional:

(20) The number of objects that are not self-identical $= 0$ if and only if $\forall x \neg(x \neq x)$.

Admittedly, we know a priori that the right-hand-side of (20) is true. And if we could know a priori that the biconditional, (20), is true, then we would certainly have a priori grounds for the belief that the left-hand-side of (20) is true. And hence a priori grounds for the belief that the number 0 exists. But now, do we know a priori that the biconditional (20) is true? I think not, and hence I think that, even though we know a priori that there are no objects that are not self-identical, we have no a priori grounds for the belief that the number 0 exists.

It is not, of course, that I seriously doubt (or think you should doubt) that the number 0 exists. Rather, the point I want to make is that, whatever else the source of our confidence in the existence of the number 0 may be, it is not simply the fact that there are no objects that are not self-identical.

But even if the line of argument given here doesn't provide us with a priori grounds for the belief in the necessary existence of numbers, it might plausibly be thought to provide us with a priori grounds for the next best claim to the necessary existence of numbers, namely, that if there are numbers, then they are necessary existents:

(21) Numbers exist $\rightarrow \Box$ Numbers exist.

Notice that for all that (21) implies, it might still be the case that there are no numbers.

Matters, however, are more complicated. For a little reflection on an exceedingly familiar argument will convince us that:

(22) $\Diamond$ Numbers exist $\rightarrow$ Numbers exist.

Suppose there is a possible world w in which numbers exist. Then, by arguments like the ones given in the preceding section, one might be able to establish that they are not identical with objects which exist contingently, and hence that they are necessary existents. Since they are necessary existents, they exist in every possible world, and, in particular, they exist in the actual world.[16]

This is, as I say, a familiar argument, as one may recognize in it the form of Anselm's argument from the possibility of existence of God to God's actual existence. One remarkable consequence of this argument is that in order for one to reject the existence of the numbers, one must be committed not only to the claim that there are no numbers, but also to the further claim that their existence is impossible, which may seem to put the nominalist in dialectical disadvantage. The argument, however, cuts both ways: in order for one to coherently maintain that there are numbers, one is forced to deny not only that numbers don't exist, but also that they could have failed to exist.

[16]The argument takes place in S5, but it is, I think, reasonable to suppose that what is metaphysically necessary in one possible world is necessary in all possible worlds.

The second remark I want to make is that the aim of the arguments I have presented is to narrow down the range of candidates to be the referents of the numerals, but not to challenge the inscrutability of numerical reference, the moral of the permutation argument with which we started, that is, the conclusion that the reference of the numerals is hopelessly inscrutable. For to admit that our use of the numerals in counterfactual contexts determines that they refer to necessary existents is to advance little if at all in settling the question to which, if one, of several necessary existents numerals refer: if we are uncertain whether numbers are Zermelo, rather than, for example, von Neumann ordinals, then, if, as most philosophers seem to suppose, pure sets are necessary existents, it will be of no help at all to know that numbers are necessary existents. In other words, even after we have corrected the thesis that the reference of the numerals is determined no more precisely than up to isomorphism, the fundamental problem of the inscrutability of reference still confronts us: no combination of intentions, central features of usage, and other non-linguistic facts that will ever enable us to discern the referents of the numerals from among the members of countless isomorphic copies of the natural number system consisting exclusively of abstracta.

I should like to conclude this section with a positive note. The moral of the permutation argument with which we started is that it is possible to permute the referents of the numerals without shift in the truth conditions of complete sentences. It may be of interest to notice that modal and counterfactual considerations of the sort we have used can be utilized to establish that it is not possible to permute the referents of the numerals with items other than numbers, and to nevertheless preserve the truth conditions, indeed the truth values, of all complete sentences of the language.

To the best of my knowledge, the first use of a permutation argument to draw a perfectly general point about reference is due to Richard Jeffrey, who envisaged a permutation of the domain of all individuals which exchanged each person with his or her social security number, but which permuted all the rest of individuals onto themselves. He then noticed that, provided that all the necessary compensatory changes are made in the interpretation of the rest of predicates of the language, such a permutation could be shown to preserve the truth values of all complete sentences of the language. Putnam's twist on the permutation argument is well-known: not only can the truth values of complete sentences be preserved, even their truth conditions can be preserved. In particular, a permutation of the domain which exchanges numbers with persons ought to preserve the truth conditions of all sentences of the language.

What I would like to suggest is that, contrary to what Jeffrey's, and, more generally, Putnam's permutation argument might have suggested, it cannot be done. To vary Jeffrey's example, suppose that we want to devise a reinterpretation of the language that preserves the truth conditions of complete sentences but exchanges the referent of the uniform numbers of the 1998 Red Sox with the members of the 1998 Red Sox. Putnam's general method consists first in specifying a permutation of the domain of individuals that exchanges each member of the 1998 Red Sox with his uniform number. For the sake of simplicity, consider the permutation of the domain, π, which exchanges each member of the 1998 Red Sox with his uniform number, but permutes all the rest of individuals onto themselves.

With π in place, we are in a position to generate a systematic reinterpretation of the language. First, reinterpret each term of the language to refer to the object onto which the permutation takes the actual referent. Since the permutation takes the number 42 onto Mo Vaughn, the numeral '42' is to be reinterpreted to refer to Mo Vaughn. Conversely, since the permutation takes the number Mo Vaughn onto the number 42, the name 'Mo Vaughn' is to be reinterpreted to refer to the number 42. The next step is to reinterpret each n-adic predicate of the language to take as its new extension the set of n-tuples whose members are the n-tuples onto which the permutation takes the n-tuples in its actual extension. Thus, for example, reinterpret the predicate 'x is even' is to be reinterpreted as: '$\pi(x)$ is even', which of course amounts to:

> 'x is even and is distinct from the number of a member of the 1998 Red Sox or is a member of the 1998 Red Sox with an even uniform number'

And, likewise, a predicate like 'x used to play first base for the Red Sox in 1998' as: '$\pi(x)$ used to play first base for the Red Sox in 1998', which again amounts to:

> 'x used to play first base for the Red Sox in 1998 and is distinct from Mo Vaughn or is the number 42'.

It is evident that such a reinterpretation will preserve the truth values of every extensional eternal sentence of the language, but the scope of Putnam's argument is supposed to be considerably wider than that. In the Appendix to (Putnam, 1981), Putnam proved that such an uniform reinterpretation is bounded to preserve the truth values of a wide range of modal and counterfactual sentences, too. For consider π', a permutation of the domain of individuals in each possible world which coincides with π in the values it assigns to actual individuals, and permutes individuals in other possible worlds onto themselves. A little reflection shows that, since the switch preserves the truth values of each atomic sentence in each possible world, it does not disturb the truth values assigned to compound sentences by modal sentential logic or by the standard sentential logic of counterfactuals.

A permutation of the domain of individuals in each possible world that exchanges actual individuals but permutes individuals in other possible worlds onto themselves will perhaps do well with modal sentential logic and with the sentential logic of counterfactuals, but it will fail to preserve the truth values of more complex modal and counterfactual sentences. For example, consider a modal sentence like:

(23) There is a number which might have played first base for the 1999 Red Sox.

The difficulty is that, under the reinterpretation induced by π', (23) is true in the actual world. For, under the reinterpretation, Mo Vaughn is both an individual to which the predicate 'is a number' applies in the actual world, and an individual to which the predicate 'plays first base for the 1999 Red Sox' applies in a possible world other than the actual world. It transpires that Mo Vaughn is, under the reinterpretation, an individual to which the modal predicate 'might have played for the 1999 Red Sox' applies in the actual world. But (23) is then a sentence whose truth value is not preserved by the systematic reinterpretation induced by π'.

It is not difficult to amend Putnam's method in order to deal with the difficulty just now raised. Take a permutation of the domain of all individuals in the actual world, and let it work across all possible worlds. That permutation will induce a systematic reinterpretation of the language that falsifies sentence (23), since, under the reinterpretation, the predicate 'plays first base for the 1999 Red Sox' will now apply to Mo Vaughn in no possible world. Yet, there remains a more persistent difficulty raised by modal and counterfactual sentences similar to the ones considered in former sections. For consider the sentence:

(24) There is a number which would have not existed if no one had survived WWII.[17]

For again, let w a possible world as much like the actual world as possible in which no one survives WWII. Mo Vaughn doesn't exist at w, and, since Mo Vaughn is both an individual to which, under the reinterpretation, the predicate 'is a number' applies in the actual world and an individual which wouldn't have existed if no one had survived WWII, under the reinterpretation, sentence (24) is evaluated true in the actual world. Of course since we all view (24) as false, we conclude that the reinterpretation fails to preserve the conditions under which we assent to modal and counterfactual sentences of the language.

A perfectly general point can be extracted from the example. To guarantee that a permutation preserves the truth-conditions for all modal and counterfactual sentences, we need to make sure that that each number is permuted onto another individual which exists necessarily, too. Otherwise, it will not be difficult to come up with counterfactuals of the form of (24) whose truth values are not preserved by the permutation. This is not a serious constraint, as it permits one to exchange the natural numbers with a wide variety of necessary existents.

Of course, the example cuts both ways: not only does it suggest that no permutation of the domain of all individuals can be used to devise an interpretation of the numerals that takes them to refer to ordinary objects, it also suggests that no permutation of the domain of all individuals induces a reinterpretation of the language that takes names of ordinary objects to refer to numbers, or to other necessary existents. Indeed, it would seem that in order to obtain a systematic reinterpretation of the language that does preserve the truth values of all modal and counterfactual sentences, one needs to make sure that each individual is exchanged with another individual which exists in exactly the same possible worlds. For it is not difficult to see that if one tries to exchange Mo Vaughn with another ordinary object which exists in different possible worlds than him, such as for example the Eiffel Tower, one will not be able to preserve the truth value of counterfactuals such as:

[17]This example is similar to one given by Vann McGee in "Does "Refers in the Language you Speak" Refer in the Language you Speak?"

(25) There is a person who would have never existed if Gustave Eiffel had never built a tower in Paris.

This may not seem a serious constraint for necessary existents such as numbers, but it is indeed a serious one for ordinary objects like Mo Vaughn; it allows us to exchange him with his unit set, but not with objects that exist in possible worlds in which he doesn't exist. These considerations suggest that the truth conditions of complete sentences would seem to determine the reference of our terms and predicates much more determinately than up to isomorphism. What is the force of these considerations is, however, a difficult question I cannot hope to address here.

1.4 Conclusion

I have argued for the the thesis that the truth of a variety of numerical sentences whose acceptance plays a crucial role in applications determines the reference of the numerals, as we ordinarily use them, much more precisely than up to isomorphism. In particular, there are a variety of modal and counterfactual sentences that play an important role in cardinal applications whose truth is incompatible with the assumption that contingent existents are candidates to be the referents of the numerals, as we ordinarily use them. An immediate consequence of this thesis is that the Aristotelian view with respect to arithmetic is not tenable. There are aspects of the natural numbers other than their position in the natural number system that matter to us. In particular, they need to exhibit a variety of features traditionally ascribed to abstracta in order for a variety of assertions made in contexts of application to be sustained.

Apart from the constraint that numbers be abstract, there are no aspects of the natural numbers that can be used to discern them from the members of another isomorphic copy of the natural system consisting exclusively of abstracta. Thus, although the problem of the inscrutability of numerical reference still confronts us, we can now venture an attractive account of the truth conditions of numerical sentences: a sentence of arithmetic, for example, is true if and only if it is true in every isomorphic copy of the natural number system consisting exclusively of abstracta. Or, equivalently, if and only if it is true in some such isomorphic copy of the natural number system.

I should like to conclude with a somewhat speculative note. Though this is not the place to discuss the question whether the inscrutability of reference is a perfectly general phenomenon, and not one restricted to the numerical case, provided that, whatever the subject matter, reference is inscrutable, a little reflection on the numerical case suggests that the claim that reference is never determined more precisely than up to isomorphism may, in the end, be much more difficult to sustain than some philosophers have advanced. For it is arguable that the numerical case provides the best possible scenario for that claim, and that, since the truth conditions of modal and counterfactual sentences defeat the claim that reference is determined no more precisely than up to isomorphism, the truth conditions of modal and counterfactual sentences can likewise determine the reference of our terms more precisely than up to isomorphism in the general case, too. This, however, is a topic for another occasion.

References

[1] Benacerraf, P. 1965. What numbers could not be. *Philosophical Review*, 74, 495-512.

[2] Benacerraf, P. 1996. Recantation *or* Any old ω-sequence would do after all. *Philosophia Mathematica*, 4, 184-189.

[3] Cartwright, R. 1987. *Philosophical Essays.* Cambridge: The MIT Press.

[4] Davidson, D. 1979. The inscrutability of reference. *Southwestern Journal of Philosophy*, 10, 227-241.

[5] Frege, G. 1884. *Die Grundlagen der Arithmetik.* Breslau: Wilhelm Koebner. English translation by J.L. Austin, *The Foundations of Arithmetic.* Oxford: Basil Blackwell, 1950.

[6] Jeffrey, R. 1964. Review of Hilary Putnam's What theories are not. *Journal of Philosophy*, 61, 79-88.

[7] McGee, V. 1993. A Semantic Conception of Truth? *Philosophical Topics.* 21, 83-111.

[8] McGee, V. Does "Refers in the Language You Speak" Refer in the Language You Speak?" Unpublished manuscript.

[9] Putnam, H. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.

[10] Tennant, N. 1997. On the necessary existence of numbers. *Nous*, 31, 307-336.

Chapter 2

Modern Set Theory and Replacement

2.1 Introduction

In "Must we believe in set theory," an article in his recent collection *Logic, Logic and Logic*,[1] George Boolos records the uncontroversial fact that it is a theorem of Zermelo-Fraenkel set theory plus the axiom of choice (ZFC) that there is a cardinal κ which is equal to $\aleph_\kappa$. That cardinal is the limit of the sequence $\{\aleph_0, \aleph_{\aleph_0}, \aleph_{\aleph_{\aleph_0}}, ...\}$, i.e., $\bigcup\{f(i) : i \in \omega\}$, where $f(0) = \aleph_0$ and $f(i+1) = \aleph_{f(i)}$. Since, as usual, cardinals are ordinals and ordinals are von Neumann ordinals, it is a consequence of standard set theory that there are at least κ sets in existence. Yet, Boolos regards the cardinal κ as an unbelievably large number, so large that it calls into question the truth of a theory according to which there are at least κ objects in existence. He thus intimates that it is perhaps more comfortable to refuse to accept the conjunction of those axioms of ZFC which entail the existence of κ than to accept the existence of κ.[2]

[1](Boolos, 1998). The essay "Must we believe in set theory?" is on pp.120-132.

[2]For those concerned by the fact that it is consistent with first-order ZFC that $2^{\aleph_0} = \kappa^+$, it may be better to pick a cardinal λ which is equal to $\beth_\lambda$. That cardinal is the limit of the sequence $\{\beth_0, \beth_{\beth_0}, \beth_{\beth_{\beth_0}}, ...\}$, and may seem unbelievably large even by the lights of someone who regards the existence of the set of real numbers as uncontroversial. The cardinals $\beth_\alpha$ are defined as usual: $\beth_0 = \aleph_0$, $\beth_{\alpha+1} = 2^{\beth_\alpha}$, and $\beth_\lambda = \bigcup\{\beth_\beta : \beta < \lambda\}$.

What Boolos contemplates in "Must we believe in set theory?" is of course not the rejection of set theory, but merely the rejection of that part of set theory that is far removed from ordinary experience, the rest of physical science and ordinary mathematics. Now: it is not obvious that Boolos' concern supplies us with a reason to refuse to accept a specific axiom (or list of axioms) of ZFC, as opposed to the conjunction of those axioms that are used in the derivation of the existence of κ. A standard pattern of argument for the existence of κ is to define a map from ω, whose existence is a consequence of the axiom of infinity, onto the sets in the sequence $\aleph_0$, $\aleph_{\aleph_0}$, $\aleph_{\aleph_{\aleph_0}}$, ..., which, by an instance of replacement, form a set, and, then, to appeal to the union axiom to derive the existence of $\bigcup\{\aleph_0, \aleph_{\aleph_0}, \aleph_{\aleph_{\aleph_0}}, \ldots\}$, which is just κ. Perhaps it is unreasonable to think that the axioms of infinity and union are particularly obvious, or self-evident, but they seem to be necessary for the development of a reasonable amount of set theory. In contrast with them, the axiom of replacement has been questioned by a number of authors, who have adduced independent reasons to doubt that replacement has received adequate justification.

The axioms of replacement give partial expression to the principle that if x is a set and R is a functional relation which associates at most one object R_y to each element y of x, then there is a set whose elements are the R_y's. A picture of the axiom is that if we replace each element y of a set x by the object associated with y by R, then the result is a set, too. This principle, independently proposed by Abraham Fraenkel and Thoralf Skolem in the 1920's, is commonly thought to be dispensable for the development of a reasonable amount of set theory, sufficiently reasonable to provide a foundation for mainstream core mathematics.

A common complaint against the axioms of replacement to be distinguished from Boolos' concern is that, though they are undoubtedly required for the development of higher set theory, they are rarely used within ordinary mathematics. The axioms of Zermelo set theory plus the axiom of choice (ZC), a theory whose axioms are all of the axioms of ZFC with the exception of the axioms of replacement, are generally assumed to be sufficient for the development of ordinary mathematics. In fact, there is not one discussion of replacement that does not call attention to the remarkable fact that no single application of replacement within ordinary mathematics could be isolated before 1971, when Harvey Friedman proved that an application of replacement on an uncountable set was required to establish that every Borel game is determined, a result that Donald Martin would later prove in 1975.[3] This fact is invariably supposed to illustrate how recondite the applications of replacement within ordinary mathematics are.

There is yet another complaint sometimes raised against replacement. This is the observation that, whereas there is a single conception of set, the iterative conception, that can be used to motivate all the axioms of ZC (with the possible exception of the axiom of choice), there is no conception of set that can be used to show the axioms of ZFC, the theory that results when replacement is added to the axioms of Zermelo set theory, to be more than an ad hoc list of principles chosen for their apparent consistency and their ability to deliver desired theorems of ordinary mathematics. The iterative conception is the view that sets are formed in stages of a certain cumulative hierarchy; a condensed version of the conception consists of principles concerning stages and principles concerning sets and stages. The principles concerning stages are aimed to ensure both that stages are well-ordered by a relation, *earlier than*, and that there is *at least* one *limit* stage, a stage that is later than some stage, but not immediately later than a stage earlier than it. The principles concerning sets and stages are designed to make sure both that a set is formed at a stage if and only if its elements are all formed at stages earlier than it and that, given some sets formed at stages earlier than some stage, they are the members of some set.

One reason the iterative conception cannot be utilized to motivate all the axioms of ZFC is that, as it has been persuasively argued, the axioms of replacement cannot be derived just from the principles concerning sets and stages implicit in the presentation of the iterative conception just now given.[4] For it is consistent with that presentation that there be a set x

[3](Martin, 1975) contains the first proof of Borel determinacy. (Friedman, 1971) showed that replacement is required for the proof of Borel determinacy, and it contains a discussion of the claim, often made prior to 1975, that replacement plays no role within ordinary mathematics.

[4]The details of this characterization of the iterative conception and its relation with replacement are presented in detail in (Boolos, 1989). Additional presentations of the iterative conception that omit replacement can be found in (van Aken, 1988) and (Potter, 1990). I should mention, though, that there are both more expansive elaborations of the iterative conception on which replacement and much more can be derived, such as Gödel's own elaboration of the iterative conception, and less expansive accounts of the iterative conception on which not even the axiom schema of separation can be derived as a consequence, such as Lavine's account in (Lavine, 1994).

and a functional relation R such that the sets R_y, which the functional relation R associates to the members y of x, occur at arbitrarily high stages of the cumulative hierarchy — this would surely be the case, for example, if the cumulative hierarchy consisted of a countable number of stages.

There is, to be sure, a different conception of set that can be used to motivate replacement, the limitation of size conception of set. According to the limitation of size idea, objects form a set if and only if there are not too many of them. But then, if x is a set, then there are not too many members of x, and consequently, if R is a functional relation, then there are not too many objects R_y associated to the members y of x by R. Different answers to the question how many are too many result in different versions of the limitation view. What is perhaps the strongest limitation of size principle is due to John von Neumann. According to von Neumann's version of the limitation of size doctrine, objects form a set if and only if they are not in one-one correspondence with all the objects there are. This yields replacement, too. If x is a set, then its members are not in one-one correspondence with all the objects there are. Nor are the objects R_y, which the functional relation associates to the members y of x, in one-one correspondence with all the objects there are.[5] Though limitation of size accounts for a large part of set theory, it also omits much of importance; in particular, it lacks the resources to account for two axioms that are crucial to any reasonable development of set theory, i.e., the axioms of power set and infinity.[6]

There are other heuristic principles that can be utilized to motivate replacement. Most set theorists, for example, regard it as plausible that, for each structural property of the universe, there is a set-sized model, indeed one of the form $\langle V_\kappa, \in \cap (V_\kappa \times V_\kappa)\rangle$, that exhibits that property. The principle of reflection is a heuristic principle that can be used to motivate replacement; indeed, it is a result due to Azriel Levy that, to the limited extent to which reflection can be formalized in the language of first-order set theory, it is equivalent to the

[5] Modulo some form of choice, which is required to infer that the range of R on x is indeed in one-one correspondence with x. Thanks are due here to Richard Cartwright.

[6] A thorough discussion of the shortcomings of different versions of limitation of size as providing a justification of both axioms can be found in (Hallett, 1984), especially chapters 4 and 8.

combination of the axioms of infinity and replacement.[7] But the point still remains that it would be a mistake to suppose that there is a set of persuasive intuitive considerations in favor of the axioms of Zermelo set theory that can be used to motivate the axioms of replacement as well.

In practice, set theorists have more argument for the adoption of replacement than the intuitive plausibility arguments I have mentioned thus far. In the context of standard set theory, the axiom of replacement is adopted because it is required for the development of an attractive theory of ordinals, on which ordinals are transitive sets well-ordered by $\in$, as well as for a justification of the much used method of definition by transfinite recursion on the ordinals. The method of transfinite recursion on the ordinals permits us to define an important cumulative hierarchy of levels or stages:

$$V_0 = \emptyset;\ V_{\alpha+1} = \mathcal{P}(V_\alpha);\ V_\lambda = \bigcup_{\beta<\lambda}, \text{ for limit ordinals } \lambda,$$

where $V = \bigcup_\alpha V_\alpha$. The interest of this cumulative hierarchy is that it gives expression to the modern, cumulative view of the set-theoretic universe, and can be utilized to motivate the axioms of standard set theory as a description of the V_α's. Thus, a little reflection on the axioms of Zermelo-Fraenkel set theory (ZFC) shows that V_ω, the first transfinite level of the hierarchy, is a model of all the axioms of ZFC with the exception of the axiom of infinity. And, in general, one finds that if κ is a strongly inaccessible ordinal, then V_κ is a model of all of the axioms of ZFC.[8] (For all these models, we take $\in$ to be the standard element-set relation restricted to the members of the domain). Now, similarly, if V_λ is an initial segment of the cumulative hierarchy indexed by a limit ordinal $\lambda > \omega$, then $\langle V_\lambda, \in \cap (V_\lambda \times V_\lambda)\rangle$ is a model of Zermelo set theory plus choice.

One may be tempted to view the axioms of Zermelo set theory plus choice as an implicit description of the initial segments of the cumulative hierarchy indexed by a limit ordinal $\lambda > \omega$ and replacement as a mere closure postulate on the ordinal levels of the hierarchy,

---

[7]Cf. (Levy, 1960). Levy established that the principle:
$\forall\alpha\exists\beta > \alpha\forall x_1, ..., \forall x_n \in V_\beta(\phi(x_1, ..., x_n)^{V_\beta} \leftrightarrow \phi(x_1, ..., x_n))$,
is equivalent to the combination of infinity and replacement.

[8]An ordinal κ is *strongly inaccessible* if and only if $\kappa > \omega$ and κ is regular and a strong limit, that is, if $\lambda < \kappa$, then $2^\lambda < \kappa$.

necessary to ensure that the cumulation of sets does not close off at $V_{\omega+\omega}$, the second limit level of the cumulative structure, but dispensable within the first $\omega + \omega$ levels of the hierarchy. Therefore, since $V_{\omega+\omega}$ contains isomorphic copies of the real and complex numbers, subsets and functions on the real numbers, and the rest of objects studied in classical mathematics, replacement would seem dispensable for set theory to provide a foundation for mathematics.[9] Indeed, it is precisely this picture of replacement, combined perhaps with the fact that it may be argued that it is independent from the iterative conception of set, that has led a number of theorists to question the assumption that there are compelling reasons to accept the axiom of replacement in the first place. Thus for example, Michael Potter omits the axiom of replacement from his development of set theory in (Potter, 1990) on the grounds that it is an "unnecessary assumption," one which is "not needed in any but the most esoteric parts of mathematics: only one result outside set theory — the assertion that every Borel game is determined — has so far been shown to require anything approaching the strength of replacement for its proof."[10]

What I would like to do in the first part of this chapter is to challenge this picture of replacement as an axiom with no applications within the first $V_{\omega+\omega}$ stages of the cumulative hierarchy. Thus, it will transpire that there are important, often neglected applications of a restricted version of replacement that take place at remarkably low levels of the cumulative hierarchy. Most of these applications will concern the special case of replacement according to which if x is a *countable* set, and R is a functional relation, then the objects, R_y's, which R associates to each element y of x, form a (countable) set.[11] Though these applications of countable replacement will not directly support the acceptance of the common, unrestricted version of the axiom, they will certainly make it plain that it would be ill-advised to abandon the axiom in the absence of another principle to replace it.

Even after the picture of replacement as an axiom far removed from the lower levels

[9]The introduction of (Friedman, 1971) outlines just this picture of replacement and its role in mathematical practice.

[10]Cf. (Potter, 1990), p. 64.

[11]Note that this restriction of replacement to countable sets coheres well with an unexceptionable and well-established part of mathematical practice, since the formation of arbitrary countable sequences in routine in modern analysis.

of the cumulative hierarchy is corrected, the question remains of whether we must believe in the unrestricted form of replacement that permits the formation of uncountable sets by replacement on other initial uncountable sets, when it is a principle with very few rare and exotic applications outside higher set theory. I shall suggest that there is no reason to think that there are persuasive direct intuitive considerations in favor of replacement that justify the axiom, and I shall then investigate the extent to which the internal needs of set theory require the full exercise of replacement. The results of this inquiry may not, in the end, decide the question of whether we must believe in replacement, but they may certainly be of help to someone with revisionary inclinations to discern what is involved in the decision to abandon replacement.

2.2 Basic set theory and replacement

I have explained that it is not uncommon for discussions of the axiom of replacement to focus on the question of what is the motivation for accepting the axiom of replacement, when replacement is a principle with such rare and exotic applications. The first point to be noticed is that this is a pseudo-question. There are important, often neglected applications of the axiom of replacement that are required within the first $\omega + \omega$ levels of the cumulative hierarchy, and are thus quite independent from the internal needs of higher set theory. Most of these applications involve countable instances of replacement, or, more precisely, instances of replacement on countable sets, and one may justifiably doubt that they provide us with compelling reasons for accepting full replacement. What they do establish is that instances of replacement on countable sets pervade set theory.

The theory that results when the axioms of replacement are omitted from the axioms of standard set theory, Zermelo-Fraenkel set theory plus the axiom of choice, is Zermelo set theory plus choice, a direct descendant of Zermelo's 1908 axiomatization of set theory. Zermelo's 1908 axiom system consisted of seven axioms: axioms of extensionality (axiom I); null set and pairs (axiom II); Aussonderungsaxioms, power set, and union (axioms III, IV, and V); an axiom of choice (axiom VI); and an axiom of infinity (axiom VII). This last axiom asserts the existence of Z_0, the Zermelo number sequence, the $\subseteq$-least set obtained from the null set by repeated application of the unit set operation.[12]

From the point of view of standard set theory, one serious omission from Zermelo's 1908 axiomatization is that it allows for the existence of non-well-founded sets.[13] For example, it is possible to construct a model of the Zermelo axioms in which there is a set whose sole member is itself. To exclude this and related anomalies, we add the axiom of regularity or foundation,

$$\textit{Reg} \quad \forall x(\exists y(y \in x) \rightarrow \exists y(y \in x \wedge y \cap x = \emptyset)),$$

to the rest of axioms of standard set theory. According to prevailing usage, Zermelo set theory is a theory which consists of versions of all of the axioms of Zermelo's 1908 system with the exception of choice and with the addition of the axiom of regularity, with the proviso that the adoption of different versions of these axioms result in different variants of Zermelo set theory.

We are going to see that what are perhaps the most common variants of Zermelo set theory are subject to a number of serious drawbacks; it will transpire, for example, that differing formulations of the axiom of infinity give rise to different versions of Zermelo set theory which are not adequate to prove the existence of sets that are remarkably simple both in terms of cardinality and in terms of their place in the cumulative hierarchy. Other shortcomings of different versions of Zermelo set theory concern their inability to prove all instances of the (first-order) principle of $\in$-induction as well as to describe and assert the first-order content of the cumulative picture. The results of this section will not depend upon the presence or absence of the axiom choice, and, consequently, there will be no loss of generality in following prevailing usage by not including choice as one of the axioms of the different variants of Zermelo set theory we will consider.

Axioms of infinity for Zermelo set theory. There are a variety of alternative formulations of the axiom of infinity, not all of them interderivable. The purpose of this section is to review the relative strength of familiar versions of infinity, and establish the inability

[12] This development occurs in (Zermelo, 1908b).

[13] The other serious omission from the point of view of standard set theory is replacement.

of some of these formulations, modulo the rest of the Zermelo axioms (Z^-), to prove the existence of HF, the set of all hereditarily finite sets, as a consequence.

Though our primary interest in this section is to discern what can and cannot be proved from first-order variants of Zermelo set theory in the absence of replacement, most of the constructions used in the proofs of the independence results of this and next sections sustain a much more general conclusion to the effect that second-order variants of Zermelo set theory can be satisfied in models in which certain first-order sentences are false. To attain further generality, then, we will focus directly on second-order formulations of Zermelo set theory.

Zermelo's original axiom of infinity asserts the existence of a set which contains the null set and which contains a unit set of any set it contains:

Inf_Z $\exists y(\emptyset \in y \wedge \forall x(x \in y \rightarrow \{x\} \in y))$.

This axiom permits one to prove the existence of the set $Z_0 = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, ...\}$, Zermelo's number sequence, as an immediate consequence, and still occurs in some presentations of standard set theory. Z^-+*Inf_Z* is the version of Zermelo set theory whose axiom of infinity is *Inf_Z*.

A more standard formulation of the axiom of infinity is:

Inf $\exists y(\emptyset \in y \wedge \forall x(x \in y \rightarrow x \cup \{x\} \in y))$.

Inf delivers the existence of ω, the first transfinite ordinal, and appears in most treatments of set theory. In what follows, I will abbreviate Z^-+*Inf* as Z, in accordance with the fact that the name Zermelo set theory is most commonly used to refer to Z^-+*Inf*

The following sentence is an ostensibly weaker axiom of infinity:

InfDed $\exists y \exists f \exists x(\mathrm{Fnc}f \wedge x \in y \wedge f : y \rightarrow_{(1-1)} y - \{x\})$.

Not only does *InfDed* fail to imply either *Inf* or *Inf_Z* (modulo the axioms of Z^-, of course), as we will see in a moment, it can even be showed that no infinite set is a member of all the models of second-order Z^-+*InfDed*.

InfDed is equivalent, modulo the axioms of Z^-, to the assertion that there exists an ordinary infinite set, a set y which cannot be put in one-one correspondence with any set of natural numbers less than some natural number n. This result is due to Russell who proved that the power set $\mathcal{P}(\mathcal{P}(x))$ of the power set $\mathcal{P}(x)$ of an infinite set x is Dedekind infinite.[14] It should be noted, however, that, absent choice, not only can it not be proved that no infinite set is Dedekind finite, it cannot even be proved that there do not exist infinite sets whose power set is Dedekind finite.[15]

The other, less common formulation of the axiom of infinity I want to consider is:

InfNew $\exists y(\emptyset \in y \wedge \forall x \forall z(x \in y \wedge z \in y \rightarrow x \cup \{z\} \in y))$.

It is evident that this axiom of infinity implies the existence of V_ω, which coincides with HF, the set of all hereditarily finite sets, as an immediate consequence, and even though it is mentioned in the second edition of (Fraenkel-Levy, 1958) and figures as the official axiom infinity in Azriel Levy's excellent text (Levy, 1973), it is seldom discussed in standard treatments of set theory.

There are, to be sure, other variations on the axiom of infinity in the literature, but I am not now concerned to present an exhaustive review. My aim rather is to point out the existence of important, and often neglected, differences among what are perhaps the most common versions of the axiom of infinity.[16]

The theories Z^-+*InfDed*, Z^-+Inf_Z, Z, and Z^-+*InfNew* having been set out, the time has come to examine dependencies among them. It is evident that every theorem of second-order Z^-+*InfDed*, Z^-+Inf_Z and Z is a theorem of second-order Z^-+*InfNew*, but one might inquire whether it is the case that every theorem of second-order Z^-+*InfNew* is a theorem of the other variations on Zermelo set theory. There is a certain set-theoretic construction

[14] If x is infinite, it can be proved that for each natural number n, the set Sn of all subsets of x of cardinality n is nonempty, and if $m \neq n$, Sm and Sn are distinct. But then, $S0$ and the function that assigns $S(n+1)$ to Sn and T itself to each subset T of $\mathcal{P}(x)$ not of the form Sn for some n bears witness to the fact that $\mathcal{P}(\mathcal{P}(x))$ is Dedekind infinite. This result is sometimes erroneously attributed to Tarski, but see (Boolos, 1994) for a detailed account.

[15] Cf. (Felgner, 1971), Chapter 3.

[16] I have suggested that many theorists overlook important differences among common versions of the axiom of infinity, but I would not want to suggest that this oversight is too pervasive. For, as I should emphasize, the relative independence, modulo the axioms of Z^-, of alternative axioms of infinity is mentioned in (Bernays, 1948), and in (Fraenkel-Bar-Hillel, 1958). And some of the drawbacks at which we shall look in the course of the discussion have been noticed before in the literature. See for example (Drake, 1974), 110-111, and (Moschovakis, 1994), Appendix B.

that will permit us to answer this question in the negative.[17] If x is a set, define the set $M_n(x)$ by the recursion:

$$M_0(x) = x,\ M_{n+1}(x) = M_n(x) \cup \bigcup M_n(x) \cup \mathcal{P}(M_n(x)).$$

Then, the basic closure of x, $M(x)$, is the union

$$M(x) = \bigcup_{n \in \omega} M_n(x)$$

If x is a (pure) transitive set, then it is routine to verify that $M_{n+1}(x)$ is just $\mathcal{P}(M_n(x))$, that $M(x)$ is a (pure) transitive which is closed under subsets, and closed under all the Zermelo operations.[18] Thus, $M(\emptyset)$ is V_ω, or, equivalently, HF, the set of all hereditarily finite sets, and, in general, $M(x)$ is the domain of the $\subseteq$-least transitive model of Z^- with the standard element-set relation which is closed under subsets and contains the set x. As a consequence, $\langle M(Z_0), \in \cap (M(Z_0) \times M(Z_0))\rangle$ and $\langle M(\omega), \in \cap (M(\omega) \times M(\omega))\rangle$ are, respectively, the $\subseteq$-least transitive models of second-order $\mathrm{Z}^- + Inf_Z$ and second-order Z with the standard element-set relation.

Lemma 1. $M(\omega) \cap M(Z_0) = HF$

Proof. That $HF \subseteq M(\omega) \cap M(Z_0)$ is an immediate consequence of the fact that both $M(\omega)$ and $M(Z_0)$ contain the null set and are closed under the power set operation.

To verify the converse inclusion, note first that $M_0(Z_0) \cap M_0(\omega) = \{\emptyset, \{\emptyset\}\}$, a member of HF. Suppose now that $M_n(Z_0) \cap M_n(\omega)$ is an element of HF. Then $M_{n+1}(Z_0) \cap M_{n+1}(\omega) = \mathcal{P}(M_n(Z_0)) \cap \mathcal{P}(M_n(\omega)) = \mathcal{P}(M_n(Z_0) \cap M_n(\omega))$. And since $M_n(Z_0) \cap M_n(\omega) \in HF$, $\mathcal{P}(M_n(Z_0) \cap M_n(\omega)) \in HF$. □

As an immediate consequence of this lemma, we obtain:

[17] The construction appears for example in (Moschovakis, 1994), p. 175. I borrow the term *basic closure* from Moschovakis.

[18] A set is *closed under subsets* if it contains every subset of each of its members. Proofs of all these basic facts can be found in (Moschovakis, 1994).

Theorem 1. *There is no $\subseteq$-least transitive model of 2nd order Z^-+InfDed with the standard element-set relation.*

Hence we conclude that there is no infinite set *whose* existence is a logical consequence of second-order Z$^-$+*InfDed*. Two other immediate consequences of the lemma are:

Theorem 2. *$\langle M(Z_0), \in \cap (M(Z_0) \times M(Z_0))\rangle$ is not a model of 2nd order Z,*

and

Theorem 3. *$\langle M(\omega), \in \cap (M(\omega) \times M(\omega))\rangle$ is not a model of 2nd order Z^-+Inf_Z.*

Proof of Theorems 2 and 3. Since $M(Z_0 \cap M(\omega) = HF$ and $Z_0 \notin HF$, $Z_0 \notin M(\omega)$. Likewise, since $\omega \notin HF$, $\omega \notin M(Z_0)$. $\square$

The dependencies established in this section may now be summarized in the following diagram:

$$
\begin{array}{ccccc}
 & & \mathrm{Z}^-+\mathit{InfNew} & & \\
 & \swarrow & & \searrow & \\
\mathrm{Z} & & \not\leftrightarrow & & \mathrm{Z}^-+\mathit{Inf}_Z \\
 & \searrow & & \swarrow & \\
 & & \mathrm{Z}^-+\mathit{InfDed} & &
\end{array}
$$

In this diagram, "$\rightarrow$" abbreviates: "is strictly stronger than," that is, "Z$^-$+*InfNew* $\rightarrow$ Z" says that every theorem of Z$^-$+*InfNew* is a theorem of Z, but that not every theorem of Z is a theorem of Z$^-$+*InfNew*. "Z $\not\leftrightarrow$ Z$^-$+Inf_Z" indicates both that there are theorems of Z that are not theorems of Z$^-$+Inf_Z and that there are theorems of Z$^-$+Inf_Z that are not theorems of Z.

One moral to be extracted from these results is that neither of what are perhaps the two most common second-order variants of Zermelo set theory has the resources necessary to guarantee the existence of sets that appear at level ω of the cumulative hierarchy, and are thus quite low down in terms of their cumulative structure — some of these sets are in fact obtainable as the range of a Δ_0 formula with domain ω, and hence minimal in terms of complexity, too.

Interpreting Z^-+*InfNew* in the theories Z^-+ *Inf*$_Z$ and Z. There is, then, an important sense in which Z^-+*InfNew* is undoubtedly superior to the two more standard variants of Zermelo set theory Z^-+*Inf*$_Z$ and Z. However, there is another question one might raise in investigating the relative strengths of Z^-+*InfNew* and the more familiar Z^-+*Inf*$_Z$ and Z: one might inquire whether they can be interpreted, or at least relatively interpreted in each other. If ϕ is a formula of the language of set theory, let $\phi^{\mathsf{M,E}}$ be the formula that results when $x \in y$ is replaced by the formula $\mathsf{E}(x, y)$ and all quantifiers are relativized to $\mathsf{M}(x)$. As usual, a *relative interpretation* of a version of Zermelo set theory, T_1, in another, T_2, consists of two formulas $\mathsf{M}(x)$ and $\mathsf{E}(x,y)$ which allow one to prove for each axiom ϕ of T_1, the sentence $\phi^{\mathsf{M,E}}$, the interpretation of ϕ, as a theorem of T_2.

Part of the interest of establishing the interpretability of a theory T_1 in another theory T_2 derives from the relative consistency result which immediately follows: If T_1 is interpretable in T_2, then proofs of $\perp$ in T_1 can be turned into proofs of $\perp$ in T_2, and thus the consistency of T_2 implies the consistency of T_1. Yet, the question whether Z^-+*InfNew* can be interpreted in Z^-+*Inf*$_Z$ and Z has an added source of interest. There is no doubt that Z^-+*InfNew* permits the development of a vast part of ordinary mathematics, but, since both Z^-+*Inf*$_Z$ and Z have revealed inadequate to to secure the existence of a vast array of subsets of V_ω, one might be inquire whether they are still adequate to formalize mathematical practice. To establish the (relative) interpretability of Z^-+*InfNew* in Z^-+*Inf*$_Z$ and Z will show that, for the purposes of formalizing mathematical practice at least, Z^-+*InfNew* is no better than the more standard variants of Zermelo set theory Z^-+*Inf*$_Z$ and Z.

We shall now see that Z^-+*InfNew*, Z and Z^-+*Inf*$_Z$, i.e., they all can be (relatively) interpreted in each other. This is of course perfectly compatible with the fact that Z^-+*InfNew* is strictly stronger than both Z and Z^-+ *Inf*$_Z$, and can be seen by reflecting on Ackermann's familiar observation that there is a model for ZF minus infinity in the natural numbers: $m \in n$ if and only if the coefficient of 2^m in the binary representation of n is 1.

Theorem 4. *Z^-+InfNew is relatively interpretable in Z.*

Proof Sketch. To produce a relative interpretation of first- and second-order Z$^-$+*InfNew* in first- and second-order Z, define the sequence M_n where $n \in \omega$ by the recursion:

$$M_0 = \omega, \ \ M_{n+1} = \mathcal{P}(M_n) - FIN(\omega),$$

with $FIN(\omega) = \{x \subseteq \omega : x \text{ is finite} \wedge \ \ x \notin \omega\}$. The proviso that $x \notin \omega$ is necessary in order to preserve extensionality. Let "M(x)" abbreviate: "$\exists n \ x \in M_n$," and construct a formula "E(x, y)" of the language of Z that expresses the relation $E(x, y)$: "either x and y are members of ω and the binary numeral for y contains a 1 at the 2^x's place, or $x \in y$ otherwise." The trick is to notice that Ackermann's coding can be extended to an isomorphism from $\langle V_{\omega+\omega}, \in \cap (V_{\omega+\omega} \times V_{\omega+\omega})\rangle$ onto $\langle M, E\rangle$. It is then routine to verify that all the interpretations of axioms of Z$^-$+*InfNew* are theorems of Z. □

An immediate corollary of this result is that Z$^-$+*Inf*$_Z$ can be interpreted in Z. And a completely parallel construction establishes both that Z$^-$+*InfNew* can be interpreted in Z$^-$+*Inf*$_Z$, and that Z itself can be interpreted in Z$^-$+*Inf*$_Z$. Thus it can be concluded that Z$^-$+*InfNew*, Z$^-$+*Inf*$_Z$ and Z are equi-interpretable.

This result provides a comforting response to the question whether Z, or Z$^-$+*Inf*$_Z$ for that matter, are still sufficient for the development of ordinary mathematics: They still are; Z$^-$+*InfNew*, a theory suited to describe an important fragment of the cumulative hierarchy, can be interpreted in both Z$^-$+*Inf*$_Z$ and Z.

Well-foundedness, cumulative structure, and replacement. One foreseeable reaction to these results is to take them merely to reveal a common oversight in standard formulations of the axiom of infinity. After all, Z$^-$+*InfNew* is a version of Zermelo set theory that proves that there is a set which contains the null set and it is closed under adjunction, $x \cup \{y\}$.

The interest of Z$^-$+*InfNew* is that it would seem to prove the existence of all sets of level $< \omega+\omega$ in the cumulative hierarchy, and thus the question immediately arises whether this theory is sufficient, when cast in second-order terms, to characterize the initial segments of the cumulative hierarchy indexed by a limit ordinal $\lambda > \omega$.

Of course a prerequisite for a (second-order) theory to characterize a class of initial segments of the cumulative structure is that it be satisfiable exclusively in models in which the element-set relation is well-founded. Since the second-order principle of set-theoretic induction is a theorem of second-order ZF, we may rest assured that second-order ZF is a candidate to characterize the initial segments of the cumulative hierarchy that are indexed by some one (uncountable) inaccessible ordinal, as in fact it does. But can we, likewise, rest assured that second-order Z^-+*InfNew* can only be satisfied in models in which the element-set relation is well-founded? We could, if we were in a position to derive the second-order principle of set-theoretic induction as a theorem of second-order Z^- + *InfNew*. Curiously, however, the answer to our question is negative. Not only can the second-order principle of set-theoretic induction not be derived from the axioms of second-order Z^-+*InfNew*, one can even make use of the Rieger-Bernays method[19] for showing the independence of the axiom of foundation to construct models of in which the element-set relation is *not* well-founded.[20]

One may be surprised to hear that that there are non-well-founded models of second-order versions of Zermelo set theory. For recall that these theories come equipped with the axiom of regularity, which is designed precisely to prevent this situation. It is of course well-known that in the context of first-order ZF, the axiom of regularity can only prevent the existence of infinite descending $\in$-chains that are first-order definable in the model. But the fault for the failure of the axiom of regularity to prevent infinite descending $\in$-chains that are not definable in the model is often supposed to lie merely in the fact that the first-order schema of replacement is ill-suited to capture the full content of this axiom.

Much less well-known is the fact that, in the absence of replacement, the axiom of regularity fails to prevent the existence of infinite descending $\in$-chains which are first-order definable. And, similarly, one may be surprised to learn that regularity fails, even in the presence of *second-order* separation, to prevent the existence of non-well-founded models of several variants of Zermelo set theory.

Theorem 5. *There are non-well-founded models of 2nd-order Z^-+InfNew.*

[19]Cf. (Bernays, 1948) and (Rieger, 1957)

[20]I am grateful to Vann McGee for asking the question of whether there are non-well-founded models of second-order variants of Zermelo set theory.

Proof. To produce a non-well-founded model $\mathcal{M}$ of 2nd order Z$^-$+*InfNew*, take the domain of $\mathcal{M}$ to be $V_{\omega+\omega}$, and let π be a permutation of the domain $V_{\omega+\omega}$ defined by:

$$\begin{aligned}
&\pi(x) = \{\{x\}\}, \text{if } x \in \{Z_0, \{Z_0\}, \{\{Z_0\}\}, ...\},\\
&\pi(Z_0 - x) = Z_0 - \bigcup\bigcup x, \text{if } x \in Z_0 - \{\emptyset, \{\emptyset\}\},\\
&\pi(Z_0 - \{\emptyset\}) = \{Z_0\}, \text{ and}\\
&\pi(x) = x \text{ otherwise.}
\end{aligned}$$

An informal, but more intuitive characterization of π is that it shifts each term forward two steps in the sequence:

$$\ldots\ , Z_0 - \{\{\{\emptyset\}\}\},\ Z_0 - \{\{\emptyset\}\},\ Z_0 - \{\emptyset\},\ Z_0,\ \{Z_0\},\ \{\{Z_0\}\},\ \{\{\{Z_0\}\}\},\ \ldots\ .$$

The relation $\in_{new} \subseteq (V_{\omega+\omega} \times V_{\omega+\omega})$ by which the symbol $\in$ is to be interpreted in $\mathcal{M}$ may then be defined by: $x \in_{new} y$ if and only if $x \in \pi(y)$. It is then immediate that $\in_{new}$ is not well-founded in $\mathcal{M}$, as Z_0, $\{Z_0\}$, $\{\{Z_0\}\}$, ... are the members of an an infinite descending $\in_{new}$-sequence in the model.

We must now see that $\mathcal{M}$ is a model of second-order Z$^-$+*InfNew*. It is routine to verify that the truth of the axioms of extensionality, null-set, pairing, infinity, and second-order separation is unaffected by π. The axioms of union, power set, and foundation require more attention and are discussed here:

Union: Let x be a member of $V_{\omega+\omega}$, and note that it is a consequence of our definition of π that $\forall x(x \in V_{\omega+\omega} \rightarrow rank(x) \leq rank(\pi(x)) \leq rank(x) + 2)$.[21] Since $\{\pi(y) : y \in \pi(x)\} \subseteq V_{rank(x)+2}$ and $V_{rank(x)+2}$ is itself a member of $V_{\omega+\omega}$, $\{\pi(y) : y \in \pi(x)\} \in V_{\omega+\omega}$. Hence we can infer that $\bigcup\{\pi(y) : y \in \pi(x)\}$ is a member of $V_{\omega+\omega}$ as well. Now, consider the set $\pi^{-1}(\bigcup\{\pi(y) : y \in \pi(x)\})$ and observe that if z is a member of $V_{\omega+\omega}$, then $z \in_{new}$ $\pi^{-1}(\bigcup\{\pi(y) : y \in \pi(x)\})$ just in case $z \in \pi(y)$ for y a member of $V_{\omega+\omega}$ such that $y \in \pi(x)$ — just in case $z \in_{new} y$ for y a member of $V_{\omega+\omega}$ such that $y \in_{new} x$.

[21] As usual, $rank(x)$, the *rank* of x, is the least ordinal α such that $x \subseteq V_\alpha$. I should emphasize that this feature of π plays an important role in the proof, for, in general, it is not the case that a one-one map of $V_{\omega+\omega}$ onto $V_{\omega+\omega}$ induces a model of Z$^-$+*InfNew*. A permutation that assigns all the finite Zermelo ordinals to their obvious counterparts in $\{Z_0, \{Z_0\}, \{\{Z_0\}\}, ...\}$ will generate a model in which union fails — consider the union of Z_0 in such a model. Compare with the Rieger-Bernays method for constructing non-well-founded models of ZF minus foundation.

Power: Suppose that x is a member of $V_{\omega+\omega}$, and note that $\pi(x)$ and $\mathcal{P}(\pi(x))$ are members of $V_{\omega+\omega}$. Observe that if $y \in \mathcal{P}(\pi(x))$, then $rank(\pi^{-1}(y)) \leq rank(y) \leq rank(\mathcal{P}(\pi(x)))$. Therefore, since $\{\pi^{-1}(y) : y \in \mathcal{P}(\pi(x))\}$ is a member of $V_{\omega+\omega}$, $\pi^{-1}(\{\pi^{-1}(y) : y \in \mathcal{P}(\pi(x))\})$ is a member of $V_{\omega+\omega}$ such that if z is a member of $V_{\omega+\omega}$, $z \in_{new} \pi^{-1}(\{\pi^{-1}(y) : y \in \mathcal{P}(\pi(x))\})$ just in case $\pi(z) \subseteq \pi(x)$ — just in case $\forall w(w \in_{new} z \rightarrow w \in_{new} x)$, as desired.

Regularity: Case 1. Suppose $x \in \{Z_0, \{Z_0\}, \{\{Z_0\}\}, ...\}$. Then, $\{x\}$ is a $\in_{new}$-minimal element of x, since $\{x\} \in_{new} x$, but $\{\{x\}\} \notin_{new} x$. Case 2. Suppose $x = Z_0 - y$ with $y \neq \emptyset$. If $y = \{\emptyset\}$, then Z_0 itself is a $\in_{new}$-minimal member of x. Else, if $y \neq \{\emptyset\}$, then $\emptyset$ is a $\in_{new}$-minimal member of x. Case 3. Otherwise, $\forall y \in V_{\omega+\omega}$ $(y \in_{new} x \leftrightarrow y \in x)$. Let y be a $\in$-minimal element of x. If $\pi(y) = y$, then y is a $\in_{new}$-minimal member of x. Else, if $\pi(y) \neq y$, then we distinguish two subcases: (a) $y \in \{Z_0, \{Z_0\}, \{\{Z_0\}\}, ...\}$. If $\{y\} \notin x$, then done. Otherwise, let $z \in x \cap \{Z_0, \{Z_0\}, \{\{Z_0\}\}, ...\}$ such that $\{z\} \notin x$ — remember that $x \in V_{\omega+\omega}$, and hence cannot contain all the elements of $\{Z_0, \{Z_0\}, \{\{Z_0\}\}, ...\}$ as members. Then, z is a $\in_{new}$-minimal element of x. (b) Else, we have that $y = Z_0 - z$ for some $z \in Z_0$ with $z \neq \emptyset$. If $z = \{\emptyset\}$, then if $Z_0 \in x$, proceed as in 1. Otherwise, $\emptyset$ itself is a $\in_{new}$-minimal member of x. □

Thus we conclude that in the absence of replacement, the usual first-order version of the axiom of regularity fails to insure that the axioms of second-order Z^-+*InfNew* are never satisfied in non-well-founded models.[22] And this result carries over to the second-order variants of Zermelo set theory discussed thus far as well. More can be said: The construction of $\langle V_{\omega+\omega}, \in_{new} \cap (V_{\omega+\omega} \times V_{\omega+\omega})\rangle$ provides us with a model of Z^-+*InfNew* which falsifies some instances of the usual first-order $\in$-induction schema:

$\in$-induction: $\exists x\phi(x) \rightarrow \exists x(\phi(x) \wedge \forall y(y \in x \rightarrow \neg\phi(y)))$,

that is: $\exists x \exists z(x = \{z\}) \rightarrow \exists x(\exists z(x = \{z\}) \wedge \forall y(y \in x \rightarrow \neg\exists z(y = \{z\}))$. Hence we conclude that, in the absence of replacement, the axiom of regularity cannot exclude the possibility of there being definable infinite descending $\in$-chains.

[22] Replacement is not the only axiom whose absence may distort the content of regularity. It is an old result of Jon Barwise that ZF$-$*Inf* cannot insure the existence of the transitive closure of a set. Part of the interest of Barwise's result is that it can readily be adapted to show that all the axioms of second-order ZF$-$*Inf* can be verified in a model in which the extension of the element-set relation is not well-founded.

Perhaps we should have anticipated these results by reflecting on the fact that the Zermelo axioms are inadequate to prove that every set has a transitive closure,[23] and this is one of the basic facts used in the standard proof that the axiom of regularity implies all the instances of the principle of $\in$-induction, modulo the rest of axioms of ZF.

One may now wonder whether Z^-+*InfNew* can be amended to prove all instances of the axiom schema of $\in$-induction, and, when cast as a second-order theory, to ensure that the universe is indeed well-founded. Two repairs suggest themselves. One repair that would achieve the desired result is to adjoin a first-order sentence to the effect that every set has a transitive closure to the rest of axioms of Z^-+*InfNew*: it is evident that first-order Z^-+*InfNew*+ "Every set has a transitive closure" proves every instance of the $\in$-schema, and, in the presence of second-order separation, enforces well-foundedness. The other option would be to adjoin a restriction of replacement to countable sets to the rest of the axioms of Z^-+*InfNew*. The result would be considerably stronger than common variants of Zermelo set theory, and it would prove that prove that every set has a transitive closure as well as all instances of the axiom schema of $\in$-induction.

There is reason to be dissatisfied with both repairs, though. The axioms of set theory are ordinarily motivated as a description of the cumulative hierarchy; the axioms of Zermelo-Fraenkel set theory enable us to define the V_α's and to prove, as a theorem, that every set x is included in some V_α. In like manner, it would be highly desirable for a variant of Zermelo set theory to be able to describe and assert the first-order content of the cumulative picture. Unfortunately, neither Z^-+*InfNew*+ "Every set has a transitive closure" nor the addition of countable replacement to Z can achieve this.

To realize this, it is sufficient to observe that there are models of second-order versions of these theories which are not isomorphic to initial segments of the cumulative hierarchy. In particular, there is a well-known construction of models of set theory which can be used to establish that there are models of second-order Z^-+*InfNew*+ "Every set has a transitive closure" which are not isomorphic to initial segments of the cumulative hierarchy: For κ an infinite cardinal, $H(\kappa)$ is the collection of all sets x whose transitive closure has only sets

[23] In (Drake, 1974), p. 110-111, Frank Drake exhibits a model of (second-order) Z in which not every set has a transitive closure.

of cardinality $< \kappa$. It is routine to verify that $H(\kappa)$, for an cardinal $\kappa > \omega$, satisfies all the axioms of (second-order) ZFC except possibly power set and replacement. For us, however, the interest of this construction is that it provides us with a recipe to construct models of Z^-+*InfNew* that are not of the form $\langle V_\lambda, \in \cap(V_\lambda \times V_\lambda)\rangle$ for a limit ordinal $\lambda > \omega$. In particular, if κ is a strong limit, then $\langle H(\kappa), \in \cap(H(\kappa) \times H(\kappa)\rangle$ is a model of (second-order) Z^-+*InfNew*+ "Every set has a transitive closure" that is not of the form $\langle V_\lambda, \in \cap(V_\lambda \times V_\lambda)\rangle$ for some limit ordinal $\lambda > \omega$.

Other models of second-order versions of Zermelo set theory that are not of the form $\langle V_\lambda, \in \cap(V_\lambda \times V_\lambda)\rangle$ for a limit ordinal $\lambda > \omega$ can be obtained merely by taking the *basic closure* of a transitive superset of V_ω.[24] Thus for example $M(V_\omega \cup \omega + \omega)$, the basic closure of $V_\omega \cup \omega + \omega$, is the domain of another model of second-order Z^-+*InfNew* which contains $\omega + \omega$, but not $V_{\omega+\omega}$ as a member. Now, to obtain a model of Zermelo set theory plus the axiom of countable replacement that is not isomorphic to an initial segment of the hierarchy, consider a model such as $\langle H(\beth_{\omega_1}), \in \cap(H(\beth_{\omega_1}) \times H(\beth_{\omega_1})\rangle$. The interest of this model is that $H(\beth_{\omega_1})$ is closed under countable subsets, and thus is a model of Zermelo set theory plus countable replacement.

The question remains whether there a variant of Zermelo set theory that is equipped with the resources to describe and assert the first-order content of the cumulative picture. The answer to this question is affirmative. There is a variation of second-order Zermelo set theory one of whose axioms explicitely asserts that sets are formed in stages. The first point to be noticed is that if we take the variables α, β, γ, ... to range over von Neumann ordinals, then the V_α's can be characterized thus:

$$x = V_\alpha \longleftrightarrow \exists f(\mathrm{Fnc}f \wedge \mathrm{Dom}(f) = \alpha + 1 \wedge \forall\beta \leq \alpha\ \forall y[y \in f(\beta) \leftrightarrow \exists\lambda < \beta\ (y \subseteq f(\lambda))] \wedge f(\alpha) = x).$$

This immediately suggests a formulation of the axiom of foundation which can be used to enforce the modern cumulative view of the set-theoretic universe. This axiom reads:

[24] As defined in section 2.

$\forall x \exists \alpha \, \exists y \; (y = V_\alpha \wedge x \subseteq y)$.

Now consider the theory that results from second-order Z when the axiom of regularity, *Reg*, is replaced by the axiom: $\forall x \exists \alpha \exists y (y = V_\alpha \wedge x \subseteq y)$. Then, the distinctions among the axioms of infinity discussed before collapse, and the axioms of second order Z $+\forall x \exists \alpha \exists y (y = V_\alpha \wedge x \subseteq y)$ do characterize the V_λ's for limit ordinals $\lambda > \omega$.

Note first that the (second-order) principle of set-theoretic induction, $\forall X \; (\exists x X x \rightarrow [\exists x X x \wedge \forall y (y \in x \rightarrow \neg X y)])$, is a theorem of the system. (Suppose Xx. Then, $\exists \alpha \; \exists x (X x \wedge x \subseteq V_\alpha)$, and, by induction on the ordinals, $\exists \beta \; (\exists x (X x \wedge x \subseteq V_\beta) \wedge \forall \lambda < \beta \; \neg \exists x (X x \wedge x \subseteq V_\lambda))$. Pick such β and x. Then of course, $\forall y \in x \; \neg X y$, since, otherwise, there would be an ordinal $\lambda < \beta$ such that $Xy \wedge y \subseteq V_\lambda$.)

Theorem 6. *$\mathcal{M}$ is a model of 2nd-order Z $+\forall x \exists \alpha \exists y (y = V_\alpha \wedge x \subseteq y)$ if and only if $\mathcal{M}$ is of the form $\langle V_\lambda, \in \cap (V_\lambda \times V_\lambda) \rangle$ for λ a limit ordinal greater than ω.*

Proof Sketch. Suppose $\mathcal{M}$ is a model of second-order Z $+\forall x \exists \alpha \exists y (y = V_\alpha \wedge \; x \subseteq y)$. By the (second-order) principle of set-theoretic induction and extensionality, the $\in$-relation of the model is well-founded and extensional, and, hence, by the Mostowski isomorphism theorem, $\mathcal{M}$ is isomorphic to a transitive $\in$-model. Without loss of generality, let us now confine attention to transitive $\in$-models of Z $+\forall x \exists \alpha \exists y (y = V_\alpha \wedge x \subseteq y)$. Suppose $\mathcal{M}$ is such a model, and let λ be the least von Neumann ordinal not in the domain. λ is a limit ordinal greater than ω, since the model satisfies the axiom of infinity and is closed under successor. Show that every member of V_λ is a member of the domain. For every $\beta < \lambda$, β is a member of the domain, and, since $\forall x \exists \alpha \exists y (y = V_\alpha \wedge x \subseteq y)$, we have that V_β itself is a member of the domain. Therefore, since the domain of M is transitive, and $V_\lambda = \bigcup \{V_\beta : \beta < \lambda\}$, we conclude that every member of V_λ is a member of the domain of $\mathcal{M}$. For the converse inclusion, observe that if x is in the model, then $V_{rank(x)}$ is a subset of the domain. But now, since λ is not in the model, neither are the V_γ's, for $\gamma > \lambda$, subsets of the domain. And thus, given that $V_{rank(x)}$ is included in the domain for each x in it, we conclude that no set of rank $> \lambda$ is in the model. □

One foreseeable source of discontent with second-order Z $+\forall x \exists \alpha \exists y (y = V_\alpha \wedge x \subseteq y)$ is that, unlike Zermelo-Fraenkel set theory, this theory enforces the cumulative hierarchy view by brute force, but it is not a natural extension of the Zermelo axioms. Doubtless, some will suggest that the moral to be extracted is that replacement may very well be the only natural principle about sets whose addition to the Zermelo axioms delivers a system of axioms which contains an implicit description of a cumulative hierarchy of levels or stages. What is the force of this consideration, however, is a question I shall not pursue here.

2.3 Must we accept (uncountable) replacement?

I think we should be admit that the results presented thus far establish very dramatically that the presupposition that replacement is rarely used in basic set theory is mistaken. This is not to deny that there remains an important question to be addressed. Since very few (albeit significant) of the applications we have considered thus far requires instances of replacement on uncountable sets (uncountable replacement), it is still legitimate to inquire what is the justification for the unrestricted version of replacement that allows for replacement on uncountable sets, when uncountable replacement is a principle with such recondite applications outside higher set theory. The remainder of this chapter is devoted to the discussion of this delicate question.

How evident is the replacement axiom?. One common response to this question is to remark that replacement, be it replacement on a countable set or on an uncountable set, is an obvious principle of set construction, an intuitive principle whose self-evidence outstrips the desirability of its consequences. I have indicated that what is probably the reason most theorists accept replacement is that it leads to an attractive theory of ordinals and permits the incorporation of transfinite recursion into set theory. Some proponents of this view would even deny that that is the primary reason the axiom is accepted; rather, set theorists accept replacement because it is a self-evident principle of set construction. In fact, it is sometimes argued that it is possible to gather some evidence for the self-evidence of replacement from a careful examination of the history of replacement. For neither Fraenkel nor Skolem, the first theorists to explicitly formulate the axiom in the context of axiomatic set theory, seemed to expect replacement to yield all these desirable consequences that now justify its adoption. And, as a consequence, one might be tempted to conceive of the history of the axiom as evidence in support of the view that, regardless of whether or not replacement is derivable from the iterative conception, it is an obvious principle whose self-evidence even outstrips the desirability of its consequences.[25] Indeed, some authors have in effect suggested that replacement emerged in the 1920s as a self-evident principle of set construction. As Shaughan Lavine explains in (Lavine, 1994):

> It did not take long for Thoralf Skolem and Abraham Fraenkel to note that Zermelo's axioms while they served Zermelo's purpose of defending his theorem were missing an important principle of Cantorian set theory — what is now the Replacement Axiom. The universal agreement that followed is remarkable, since the axiom wasn't good for anything. That is, at a time when Replacement was not known to have any consequences about anything except the properties of the higher reaches of the Cantorian infinite, it was nevertheless immediately and universally accepted as a correct principle about Cantorian sets.[26]

The interesting implied suggestion is that the axiom of replacement might have emerged as a principle of set construction that is evident on the pre-axiomatic concept of set that underlay Cantor's theory of sets and transfinite numbers. Lavine's picture promotes the impression that "even today, the self-evidence of replacement outstrips its applications. We accept it because it is true of the combinatorial notion of set, that is, because it is self-evident"[27] — an impression that informal expositions of the subject would seem to reinforce when they suggest that all the axioms of standard set theory can be justified by appeal to intuitive considerations alone.[28]

At the outset, let me acknowledge that I have no conclusive argument that will persuade a defender of Lavine's picture to abandon the view that both Fraenkel and Skolem regarded replacement as a self-evident principle of set construction. Nor do I intend to deny that the axiom of replacement enjoys a measure of obviousness or evidence, as is suggested by the fact that Cantor, Mirimanoff and Hartog had confidently used similar principles of set

[25] Two extremely useful histories of the axiom is given in (Hallett, 1984) and (Lavine, 1994).

[26] (Lavine, 1994), p. 5.

[27] Cf. (Lavine, 1994), p. 216.

[28] Lavine's remarks are motivated by the conviction that replacement is evident on a certain conception of set, the combinatorial conception. Unfortunately, a discussion of Lavine's account of the combinatorial conception of set is beyond the scope of this chapter.

construction in proofs long before the axiom was ever explicitly formulated. All I shall offer are some considerations that make it plausible to suppose that the inability of Zermelo's 1908 axiom system to sanction constructions of countable sets that seemed unobjectionable from the point of view of pre-axiomatic set theorists may have played an important role in both Fraenkel and Skolem's respective proposals of the replacement axiom. The ability to sanction such constructions is, by itself, a significant application of the axiom of replacement that is not restricted to "the higher reaches of the Cantorian infinite."

As Zermelo states the purpose of his axiomatization of set theory in (Zermelo, 1908), it appears to have been:

> to show how the entire theory created by Cantor and Dedekind can be reduced to a few definitions and seven principles, or axioms, which appear to be mutually independent.[29]

Zermelo's 1908 axiom system, recall, consisted of axioms of extensionality; null set and pairs; Aussonderungsaxioms, power and union; an axiom of choice; and an axiom of infinity asserting the existence of Z_0. One serious omission from Zermelo's axiomatization is that Zermelo's axioms do not sanction certain constructions permitted in Cantor's theory of sets and transfinite numbers. To use the standard example, Zermelo's 1908 are inadequate to establish that

$$\{Z_0, \mathcal{P}(Z_0), \mathcal{P}(\mathcal{P}(Z_0)), \ldots\}$$

is a set — where Z_0 is Zermelo's number sequence and $\mathcal{P}$ denotes the power set operation.[30] It is the axiom of replacement that permits this construction in the context of standard set theory.

However serious, the absence of this and other axioms from Zermelo's 1908 list is perhaps understandable. One of the principal motives of Zermelo's axiomatization appears to have

[29](Zermelo, 1908, p. 200).

[30]The results of the preceding section gives us ample evidence for this, and indicate that the inability to collect $\{Z_0, \mathcal{P}(Z_0), \mathcal{P}(\mathcal{P}(Z_0)), \ldots\}$ is far from being the most dramatic instance of example of a construction that cannot be carried out in the context of Zermelo's axiom system.

been to clarify the set existence principles that are used in the course of his 1904 proof that every set can be well-ordered. Not surprisingly, these are the axioms of separation, elementary sets, power set, union, and, of course, the axiom of choice.[31]

Whatever the reason for Zermelo's 1908 omission of replacement, it might be argued that, by Zermelo's own standards, the axiom fits well with the rest of axioms in Zermelo's 1908 list. One reason for this is that it is often remarked that, for Zermelo, a principle that is an unexceptionable, well-established part of mathematical practice as it is practiced elsewhere ought to be accepted in set theory as well. This is an important consideration in support of the axiom of choice in (Zermelo, 1908a), and it is not difficult to imagine a similar argument in favor of a restricted form of replacement: since countable sequences pervade modern analysis and seem routine and benign, set theory, too, ought to sanction the formation of arbitrary countable sequences. This consideration alone could be used to motivate a restricted form of replacement according to which one could form arbitrary countable sets of Zermelo sets.

Unfortunately, no similar consideration would seem sufficient to deliver the full strength of replacement.[32] For no matter how routine or benign arbitrary countable sequences may seem, arbitrary uncountable sequences are indeed a novelty and very few results within ordinary mathematics seem to depend upon their availability, witness the fact that almost 50 years had to elapse since Fraenkel and Skolem first proposed the axiom in order for a mathematician to isolate an application of replacement on an uncountable set within ordinary mathematics.[33]

The other reason one might be tempted to suppose that replacement is in line with the rest of Zermelo's axioms is that it is sometimes argued that Zermelo's axiomatization is motivated by a doctrine of limitation of size, and replacement falls rather directly out of the

[31]It is then no wonder that, as Frank Drake parenthetically remarks in (Drake, 1974), p. 114, Zermelo had been reported as saying that he *forgot* to add replacement to his 1908 axiom system; no principle like replacement is used in the course of Zermelo's 1904 proof that every set can be well-ordered.

[32]It may be of interest to notice that a perfectly analogous situation would seem to obtain with respect to the axiom of choice; all classical applications of choice in analysis can be justified on the basis of the much weaker axiom of dependent choices, and most of them can be sustained with the help of the weaker principle of countable choice.

[33]This is Friedman and Donald Martin's result that replacement is required to establish that every Borel game is determined. See note 3.

limitation of size account. Michael Hallett, for example, has claimed that it is fairly clear that Zermelo accepted the limitation of size hypothesis, and that separation gives partial expression to that view. Perhaps so, but notice that the thought expressed by separation is compatible with an interpretation of the limitation of size doctrine based on comprehension, but not on cardinality: One could identify, first, a collection of unobjectionable, safe operations to generate new sets from initial sets, and then restrict attention to those sets that can be "separated" from sets obtained by repeated application of these operations on certain initial sets, such as the null set or Z_0. Zermelo himself outlined a remarkably similar picture in (Zermelo, 1908a) when he explained that:

> if in set theory we confine ourselves to a number of established principles such as those that constitute the basis of our proof [that every set is well-ordered] — principles that enable us to form initial sets and to derive new sets from given ones — then all such contradictions can be avoided.

If we now confine ourselves to sets which can be "separated" from sets obtained from initial sets by repeated applications of the established principles to which Zermelo referred, then we will not even be able to sanction the formation of arbitrary countable sequences, much less uncountable ones. Indeed, as Hallett suggests, it is plausible to attribute this very interpretation of the limitation of size doctrine to Fraenkel in the mid 20's. Thus I think it is not clear that replacement falls out of Zermelo's 1908 picture of what sets are.

Nor did Zermelo explain what, exactly, the axioms in his 1908 list were axioms *for*. As axioms are ordinarily conceived, they may be supposed to enjoy some preferred epistemological status deriving from their accordance to some pre-axiomatic concept. One might have thought, in particular, that there is some pre-axiomatic concept of set that justifies the adoption of Zermelo's 1908 axioms. Now: Zermelo did not conceive of his axioms as axioms *for* the concept of set that underlay Cantor's theory of sets and transfinite numbers —in part because he attributed to Cantor the naive concept of set that fell prey to Russell's paradox. Instead, Zermelo made no claim to have offered an account of what sets are other than characterizing them as elements of an abstract domain structured by the element–set relation. Accordingly, it is probably best to think of the Zermelo axioms as closure conditions imposed on that abstract domain: The axioms of null set and infinity ensure that the domain of set theory contains both the null set and Z_0, the Zermelo number sequence. Other axioms require the domain to be closed under the operations of pairing $\{x, y\}$, union $\bigcup x$, and power $\mathcal{P}(x)$. Separation, for its part, guarantees that the domain contains every subset of every set it contains. Even extensionality could be conceived as a closure condition that requires the domain to contain a witness of every inequality between two sets, an object that is a member of one but not to the other.

Although Zermelo didn't claim to have offered an intuitive account of the conception of set he was assuming, he certainly wanted to claim for his axioms that they yield the theorems of the Cantorian theory of transfinite numbers. This was, after all, one of the stated aims of the axiomatization. Since, according to Zermelo, Cantor's naive conception of set had proved bankrupt, Zermelo writes:

> There is at this point nothing left for us to do but to proceed in the opposite direction and, *starting from set theory as it is historically given*, to seek out the principles required for establishing this mathematical discipline. (emphasis added)[34]

What Abraham Fraenkel and Thoralf Skolem would soon question in the 1920s is that Zermelo's 1908 axioms provided one with all the principles required to accomplish that end. Abraham Fraenkel (Fraenkel, 1921, 1922) and Thoralf Skolem (Skolem, 1922) independently observed in the 1920s that "Zermelo's axiom system is not sufficient to provide a foundation for ordinary set theory."[35] Fraenkel had already noticed in (Fraenkel, 1921) that Zermelo's seven axioms were not sufficient to prove the existence of the set $\{Z_0, \mathcal{P}(Z_0), \mathcal{P}(\mathcal{P}(Z_0)), ...\}$, and concluded that Zermelo's axiom system could not prove the existence of sets of cardinality $\geq \aleph_\omega$.[36] Skolem, for his part, raised a more general concern in (Skolem, 1922):

> If M is an arbitrary set, it cannot be proved [in Zermelo's system] that M, $\mathcal{P}$ M, $\mathcal{P}^2$ M, ... and so forth ad infinitum form a set.[37]

[34](Zermelo, 1908, p. 200).

[35](Fraenkel, 1921; 1922). The quote is from (Skolem, 1922, 292).

[36]Fraenkel, it should be noticed, might have been assuming the truth of the Generalized Continuum Hypothesis.

[37](Skolem, 1922, p. 296).

And he even sketched the construction of an inner model of the Zermelo axioms which excludes the existence of the set $\{Z_0, \mathcal{P}(Z_0), \mathcal{P}(\mathcal{P}(Z_0)), ...\}$, a set generated from the initial parameter Z_0 by repeated application of the power set operation. Skolem's model is given by:

$$\{x : \exists n \in \omega \bigcup^n x = \emptyset\} \cup \{x : \exists n \in \omega\ \exists y\ \exists m \in \omega\ (\bigcup^n x = y \wedge \bigcup^m y = \emptyset\},$$

when the predicate symbol "$\in$" is interpreted to denote the standard element-set relation, i.e., $\langle V_{\omega+\omega}, \in \cap (V_{\omega+\omega} \times V_{\omega+\omega})\rangle$.[38]

What is worth dwelling on in Skolem's complaint is that the pattern of set construction to which he is referring is perfectly general and exceedingly simple, since it consists in the formation of a countable sequence of sets from an initial parameter, Z_0, by repeated application of the power set operation. I have already noticed that the formation of arbitrary countable sequences pervades ordinary mathematics, and not just the Cantorian theory of sets and transfinite numbers, and hence the fact that Zermelo's 1908 axioms do not sanction that construction ought to be regarded as a significant loss.

Other theorists had detected a similar difficulty by 1922. In the abstract of a lecture read at a session of the 1922 meeting of the American Mathematical Society, (Lennes, 1922), Nels Johann Lennes claimed that Zermelo's system could not prove "that an arbitrary collection more than finite in number is a set." Though he published no details, he must have had examples of countable collections whose existence could not be proved by the 1908 Zermelo axioms. What this suggests is that Lennes must have called attention to the drawback of Zermelo's axiomatization just now discussed, and not to a defect concerning the system's inability to yield the existence of very large cardinals. Lennes proposed to modify Zermelo's axiomatization "so as to identify as a set any collection of objects having the same cardinal number as that of some Zermelo set," which is in effect to propose a version of replacement.

All these theorists might have believed that the set $\{Z_0, \mathcal{P}(Z_0), \mathcal{P}(\mathcal{P}(Z_0)), ...\}$ provides "the simplest instance of a set whose existence cannot be proved by means of the Zermelo axioms,"[39] but the drawbacks of Zermelo's axiomatization are much subtler. For, presum-

[38] And it is hence not identical with $\bigcup\{\omega, \mathcal{P}(\omega), \mathcal{P}(\mathcal{P}(\omega)), ...\}$, as it is sometimes erroneously said, cf. (Wang, 1970). HF is a member of the former but not a member of the latter.

[39] The quote is from (Fraenkel and Bar-Hillel, 1958).

ably unbeknownst to them, Zermelo's axiom system, and current versions of Zermelo set theory for that matter, are ill-suited to sanction a variety of apparently unproblematic set-theoretic constructions. Indeed, the results of the first part provide us with an impressive catalogue of independence results for both first- and second-order Z+Inf_Z, an immediate descendant of Zermelo's 1908 axiom system.

I have depicted Zermelo's axiom system in the worst possible light, as a collection of principles chosen for their apparent consistency and ability to sanction all the set existence assumptions that underlie Zermelo's proof that every set can be well-ordered, but which nevertheless fail to yield the existence of exceedingly simple sets, sets that occur at remarkably low levels of the cumulative hierarchy. Now: I don't wish to suggest that Fraenkel and Skolem were aware of all these difficulties, I don't think they were. All I want to suggest is that it is important to distinguish two different concerns in their respective complaints that Zermelo's axiom system is insufficient "for the foundation of legitimate set theory." Whereas Fraenkel's main source of concern, (Fraenkel 1921, 1922), appears to have certainly derived from the system's inability to secure the existence of specific large sets, sets of power $\geq \aleph_\omega$, in (Skolem, 1922), Skolem appears to have been worried about the system's inability to sanction an apparently unproblematic construction of a denumerable set from an initial parameter by repeated application of some one Zermelo operation. This fact may perhaps be obscured by the fact that Skolem's only example involves a set whose union has cardinality $\beth_\omega$, but, to insist that replacement had in the early 1920s no use other than to secure the existence of relatively high cardinals is to belittle the fact that replacement was motivated through the need for specific countable sets that seemed unobjectionable from the point of view of pre-axiomatic set theorists.

When Fraenkel asserts that "Zermelo's seven axioms are not sufficient for the foundation of legitimate set theory," he certainly seems to be assuming that a system that is sufficient for that purpose must, at the very least, provide us with the resources necessary to prove the existence of sets of power $\geq \aleph_\omega$. The reason is not difficult to ascertain: $\aleph_\omega$ is a relatively low cardinal by the lights of Cantor's theory of transfinite numbers, and hence an adequate axiomatization of set theory ought to be able to reconstruct this aspect of Cantor's theory.[40]

[40] Fraenkel's presumption, it should be noticed, is seldom challenged. Most set theorists accept the exis-

Fraenkel had noticed in (Fraenkel, 1921) that merely extending the axiom of infinity to assert the existence of the set $\{Z_0, \mathcal{P}(Z_0), \mathcal{P}(\mathcal{P}(Z_0)), ...\}$ would have rendered the resulting system vulnerable to more general counterexamples. He concluded the need for a remedy of a more general character. One year later he would propose the axiom of replacement:

> *Replacement axiom*: If M is a set and if each element of M is replaced by "a thing of the domain B" then M is transformed into a set.

Undoubtedly, Fraenkel's proposal depends, for it to be acceptable, upon a precise explanation of his picture of *replacing* the members of a set by objects of the domain. But he could have done so by making use of an adequate notion of function: If f is a "function" and m is a set, then $\{f(x) : x \in m\}$ (the range of f on m) is a set. Whether or not he was in 1922 in a position to formulate an adequate version of the axiom is a question I shall not pursue here, but I should mention that there is some evidence discussed both in Hallett and Lavine's histories of the axiom to suspect that he was not in such a position.

As for Skolem, he began his attack to Zermelo's system with the observation that "Zermelo's axiom system is not sufficient to provide a complete foundation of the *usual* theory of sets."[41] And he quickly pointed out to the system's inability to sanction the construction of a set from an initial parameter M by repeated application of the power set operation. Perhaps unbeknownst to him, Skolem's point can be strengthened considerably: it can be proved as a theorem of ZF that if M is an infinite set of level ω in the cumulative hierarchy such that $M \cap Z_0 \in HF$, then M is not a member of $\mathcal{Z}$, and thus its existence is not provable in the context of Zermelo's axioms.[42] ω is a prominent example of such a set, but

tence of higher transfinite cardinals as a matter of course, and part of the reason is that their existence can be proved in standard (Zermelo-Fraenkel) set theory; indeed, that is often supposed to be one of the virtues of the system. It is only recently, (Boolos, 1998), that someone has argued that it is a perfectly sensible view to hold both to doubt the truth of theorems concerning the higher infinite that are provable in ZF and to regard theorems concerning sufficiently low levels of the cumulative hierarchy as unquestionably true.

[41](Skolem, 1922). The emphasis is mine.

[42]Thus: M obviously is a member of

$$\mathcal{M} = \bigcup\{M, \mathcal{P}(M), \mathcal{P}(\mathcal{P}(M)), ...\},$$

but $\mathcal{M} \cap \mathcal{Z} \subseteq HF$. The latter fact is proved by a simple induction. By hypothesis, $M \cap Z_0 \in HF$. Suppose now that $\mathcal{P}^n(M) \cap \mathcal{P}^n(Z_0) \in HF$. Then, since $\mathcal{P}^{n+1}(M) \cap \mathcal{P}^{n+1}(Z_0) = \mathcal{P}(\mathcal{P}^n(M) \cap \mathcal{P}^n(Z_0))$ and notice that ω is a set which can be obtained from the null set by repeated application of the von Neumann successor operation, $x \cup \{x\}$. Likewise, it should be obvious how to use the former observation to obtain further sets which can be constructed from an initial parameter by repeated operation of some Zermelo operation but whose existence is independent from Zermelo's 1908 axiom system.

It would probably be a mistake to attribute too much generality to Skolem's remarks, but the point remains that the pattern of set construction to which he calls attention is both perfectly general and exceedingly simple. Thus, it can still be argued that he motivates replacement as a repair to a serious shortcoming of Zermelo's 1908 axiomatization that is not restricted to the higher reaches of the Cantorian infinite.

In contrast to Fraenkel's remedy, Skolem's repair to Zermelo's axiom system was quite precise and much closer to the standard schematic version of the axiom:

> In order to remove this deficiency of the axiom system we could introduce the following axiom:
> Let U be a definite proposition that holds for certain pairs (a,b) in the domain B; assume, further, that for every a there exists at most one b such that U is true. Then as a ranges over the elements of a set M_a, b ranges over all elements of a set M_b.[43]

In the context of his paper, "a definite proposition that holds for certain pairs (a,b) in the domain B" is just a well-formed formula of the language of set theory. Accordingly, he offered a general procedure for obtaining a formula sufficient to prove the existence of $\{M, \mathcal{P}(M), \mathcal{P}(\mathcal{P}(M)), ...\}$ for an arbitrary Zermelo set M.

Neither Fraenkel nor Skolem advocated the addition of replacement to Zermelo's system, neither commented on whether they conceived of replacement as a self-evident, or to a certain extent obvious principle, or even on whether they thought their remedies to systematize prior practices. This much is certain. Skolem expressed some caution in the formulation of his repair to Zermelo's axiomatization; he just claimed that "in order to remove this

$\mathcal{P}^n(M) \cap \mathcal{P}^n(Z_0) \in HF$, we have that $\mathcal{P}^{n+1}(M) \cap \mathcal{P}^{n+1}(Z_0) \in HF$, too. Compare with Lemma 1 above.

[43](Skolem, 1922).

deficiency of the axiom system *we could* introduce the following axiom."[44] Such caution suggests that their emendation to Zermelo'a axiomatization had a tentative character not to be neglected. As for Fraenkel, he would soon refer to the axiom of replacement as:

> an axiom proposed by me as a *stopgap*, an axiom which nevertheless would seem to be too sweeping to be called upon without a painstaking investigation of its necessity.[45]

Michael Hallett cites some evidence (albeit inconclusive) to suggest that Fraenkel even regarded replacement as a suspicious principle. In 1926, he described the axiom as an "unpleasant far reaching axiom," and he often referred to the "special" sets that the axiom generated as "very comprehensive sets."[46] The interest of this denomination is that it suggests that replacement may have conflicted with Fraenkel's own interpretation of the limitation of size doctrine according to which only collections that are bounded by — or subsets of — some Zermelo set can form sets.[47]

Whatever the source of Fraenkel's reluctance to accept the axiom of replacement, he did not accept the necessity of the axiom for the purposes of general set theory, and, at least up to 1958, he continued to regard the theory of ordinals and the theory of cardinals $\geq \aleph_\omega$ as "special set theory." For example, in (Fraenkel, 1927), he writes:

> ... *general* set theory can be derived in its full extent from the axioms I-VII, the Zermelo axioms.[48]

Curiously, however, (Fraenkel, 1927) contains a generalization of Zermelo's axiom of infinity which is provably equivalent to a restriction of replacement to countable sets. Roughly speaking, Fraenkel's generalization of the axiom of infinity states that for every set x and every function F, there is a set y which contains x and contains $F(z)$ whenever it contains z.

[44](Skolem, 1922). The emphasis is mine.

[45](Fraenkel, 1925), p. 251. The emphasis is mine.

[46](Fraenkel, 1926), p. 131.

[47](Hallett, 1984).

[48]Cf. (Fraenkel, 1927). p. 139. The translation is Hallett's.

The interest of this fact is that it suggests that Fraenkel may have regarded the formation of arbitrary countable sets as part of general set theory, too.[49]

Lavine's interpretation of the early history of the axiom seems to rest on the presupposition that the mere addition of sets of cardinality $\geq \aleph_\omega$, rather exotic objects, would not seem sufficient to explain the widespread acceptance of the axiom that followed Fraenkel and Skolem's respective proposals. Thus, in (Lavine, 1994), he repeats:

> Replacement received no historical justification as a systematization of prior practice or the like: the fact that its applications are so recondite shows that it had to do little with prior practice. The application of replacement came only after its acceptance. Both Fraenkel and Skolem saw that replacement was a self-evident principle concerning combinatorial collections before von Neumann discovered that the axiom was good for something and even before they were convinced of the utility of combinatorial collections.[50]

And yet, motivating the axiom of replacement as a repair designed to sanction an apparently unproblematic procedure for constructing new countable sets from an initial parameter and a Zermelo operation is significant, and renders replacement a new generator in the Zermelian setting, and, in particular, as a remedy to an important drawback of Zermelo's axiomatization that is not restricted to the "higher reaches of the Cantorian infinite."

Ordinals and replacement. I have suggested that both Fraenkel and Skolem had motivated their respective proposals of replacement as a repair designed to generate countable sets like $\{Z_0, \mathcal{P}(Z_0), \mathcal{P}(\mathcal{P}(Z_0)), ...\}$, whose existence seemed unobjectionable from the point of view of pre-axiomatic set theory. However, it was only von Neumann's discovery of his theory of ordinals and the formalization of transfinite recursion, let alone its proof, that required the full exercise of the axiom of replacement. Neither Fraenkel nor Skolem seemed to expect replacement to yield these desirable consequences, and some authors, Michael

[49](Bernays, 1942) contains a discussion of Fraenkel's variant of the axiom of infinity and its relation to the claim that every countable class is a set. An important part of Bernays' article is also devoted to show that the result of adjoining the latter claim to the Zermelo axioms is sufficient for the development of a vast part of modern analysis.

[50](Lavine, 1994), p.216.

Hallett for example, have credited John von Neumann, and not Fraenkel or Skolem, with the *discovery* of the axiom.[51]

John von Neumann published the details of his theory of ordinals in 1923, but, as a letter to Zermelo of 15 August 1923 indicates, by then, he had already developed his own axiomatization of set theory, which contained a form of replacement. As von Neumann would later comment on the system, "Fraenkel's replacement axiom is added. This (among other things) is necessary for the setting up of the theory of ordinal numbers."[52] Nevertheless, von Neumann's treatment of ordinal numbers did not depend on the special characteristics of his axiomatization, and could, according to him, be embedded in Zermelo's axiomatization provided only that this is supplemented with the axiom of replacement.

As von Neumann indicates in (von Neumann, 1923a), his theory is intended to impose a specific form on the set representatives of the ordinal numbers: "Every ordinal is the set of the ordinals that precede it." To be quite specific, von Neumann defined for each well-ordered set $\langle X, <\rangle$ a "numeration" of X to be a function f on x such that for all x in X $f(x) = \{f(y) : y < x\}$. And he then identified the ordinal number of the well-ordered set $\langle X, <\rangle$ with the range of the numeration f on X:

> If x_1, x_2, x_3, and x_4 are the 1st, 2nd, 3rd, and 4th elements of x, respectively, then clearly for every numeration $f(x)$ of X we have
> $f(x_1) = \emptyset$,
> $f(x_2) = \{\emptyset\}$,
> $f(x_3) = \{\emptyset, \{\emptyset\}\}$,
> $f(x_4) = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$
> consequently, if X has 0, 1, 2, or 3 elements, its ordinal is, respectively,
> $\emptyset$,
> $\{\emptyset\}$,
> $\{\emptyset, \{\emptyset\}\}$, or
> $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$

One important respect in which von Neumann's theory requires the full strength of replacement is that, without it, it cannot be proved that every well-ordered set has a

[51](Hallett, 1984), ch. 8.

[52](von Neumann, 1923b).

numeration: in the absence of countable replacement, it is not even possible to prove the existence of a numeration for $\langle Z_0, \in\rangle$, i. e. ω; absent a more general form of replacement, it is not even possible to prove the existence of a numeration for $\langle \mathcal{P}(Z_0), <\rangle$ where $<$ is a well-ordering of $\mathcal{P}(Z_0)$. Replacement is tailor-made for that purpose; with the aid of the replacement axiom, von Neumann could establish the existence of a numeration for each well-ordered set $\langle X, <\rangle$, and he then identified the ordinal number of a well-ordered set $\langle X, <\rangle$ with the range of a numeration f on X. In addition, he established both the existence of a unique ordinal for each well-ordered set, and the fact that isomorphic well-ordered sets receive the same (von Neumann) ordinal number. The other important respect in which von Neumann's theory of ordinals relies on the axiom of replacement is that it is replacement that is required to justify definitions by transfinite recursion, and these seem indispensable for the definition of operations of ordinal addition and multiplication in complete generality, of which more shortly. Indeed, even though von Neumann never explicitly advocated the incorporation of replacement into Zermelo's system, he nevertheless explained that "in fact, I believe that no theory of ordinals is possible at all without this [replacement] axiom."[53]

There is no question that the von Neumann theory of ordinals requires the full exercise of replacement, but this is yet a far cry from the claim that no theory of ordinals is possible in the absence of replacement. Perhaps von Neumann assumed that in order for a theory of ordinals to be satisfactory, it must be such that it assigns ordinals to well-ordered sets in a process that depends on the *set* of previous assignments to the successive initial segments of the set. But to be able to form the set of previous assignments invariably requires one form of replacement or another. For example, as ordinals are assigned to well-ordered sets in von Neumann's theory, the ordinal assigned to a well-ordered set coincides with the set of previous assignments to initial segments of the set. This requires, of course, successive availability of the set of previous assignments to the different initial segments of the well-ordered set, which can only be guaranteed by a form of replacement.

Now: it may justifiably be doubted that the requirement to form the ordinal of a well-ordered set in a process of the sort just outlined is a necessary condition for a theory of

[53](von Neumann, 1925).

ordinals to be satisfactory. Instead, it is arguable that all is required for a theory of ordinals to be adequate is that there be a definable functional relation, *ord*, that associates with each well-ordered set $\langle X, <\rangle$ a set representative, $ord(\langle X, <\rangle)$, but such that $ord(\langle X, <\rangle) = ord(\langle Y, <\rangle)$, whenever $\langle X, <\rangle$ and $\langle Y, <\rangle$ are isomorphic structures. Is replacement necessary to ensure the existence of such a definable functional relation? It may be of interest to note is that one cannot hope to *prove* the answer to be affirmative. To realize this, it is enough to recall Gödel's observation that $\langle L_{\aleph_\omega}, \in \cap (L_{\aleph_\omega} \times L_{\aleph_\omega})\rangle$ is a model of Zermelo set theory plus the existence of a definable well-ordering of the universe. For then, one can use the definable well-ordering of $L_{\aleph_\omega}$ to define a functional relation that associates with each well-ordered set $\langle X, <\rangle$ in $L_{\aleph_\omega}$ the least structure that is isomorphic to it.[54]

Not only can it not be proved that replacement is required for the existence of an adequate theory of ordinals, there is an attractive approach to the theory of ordinals due to Dana Scott that need not rest upon the availability of replacement. Scott devised a general method of definition that permitted him, for example, to rescue the Fregean treatment of cardinal numbers for Zermelo-Fraenkel set theory; in particular, he showed how to define the cardinal number of a set as the set of sets equinumerous with it of least rank. And, similarly, he showed how to define the ordinal of a well-ordered set as the set of well-ordered sets isomorphic to it of least rank.

However, Scott's technique depends upon set-theoretic facts which are ordinarily established with the aid of the axiom of replacement. For example, one ordinarily uses the replacement to justify the method of transfinite recursion on the ordinals, and then to define the V_α's, and to prove a theorem to the effect that every set is a subset of some V_α. Once this theorem is in place, it is possible to define the rank of a set, x, as the least ordinal α such that $x \subseteq V_\alpha$, and, more importantly, for the definition of the Scott ordinal of a well-ordered set $\langle X, <\rangle$ as the set of all well-ordered sets $\langle Y, <\rangle$ of least rank that are isomorphic to $\langle X, <\rangle$.

Despite appearances to the contrary, replacement is not strictly required to prove that every set is formed at some level of a cumulative hierarchy $\bigcup_\alpha V_\alpha$. For recall that the theory

[54] I am indebted to Vann McGee for this example. Other limit ordinals $\lambda > \omega$ would do as well.

Z $+\forall x \exists \alpha \exists y (y = V_\alpha \wedge x \subseteq y)$, where "$y = V_\alpha$" abbreviates:

$$\exists f(\mathrm{Fnc}(f) \wedge \mathrm{Dom}(f) = \alpha + 1 \wedge \forall \beta \leq \alpha \ \forall y[y \in f(\beta) \leftrightarrow \exists \lambda < \beta \ (y \subseteq f(\lambda))] \wedge f(\alpha) = x).$$

Now, this theory may undoubtedly seem ad hoc or unnatural as a version of Zermelo set theory, but it certainly provides us with a setting in which Scott's plan can be carried out without incident.

A more natural presentation of set theory that would serve the same purpose is due to Richard Montague and Dana Scott, who developed an axiomatization of set theory designed to enforce the cumulative picture that dispenses with replacement entirely. Instead, they started with the notion of "partial universe" as primitive and showed how a remarkably simple set of axioms for sets and partial universes is sufficient to assert the first-order content of the picture of the set-theoretic universe as a cumulative hierarchy of stages or "partial universes." Scott, in particular, moved to a two-sorted language with variables x, y, z... for sets, and variables V, V', ... for partial universes, and produced axioms governing them.[55] The axioms of Scott's theory are those of extensionality and separation, an accumulation axiom,

Accumulation: $\forall V \forall x(x \in V \leftrightarrow \ \forall V'(V' \in V \wedge x \in V' \vee x \subseteq V'))$,

which makes sure that the members of a level are the members and subsets of previous levels (the levels are ordered by the element-set relation), and an axiom of restriction,

Restriction: $\forall x \exists V \ x \subseteq V$,

which states that every set is included in some level. It can be shown that all the axioms of Zermelo set theory except for choice and infinity follow from these axioms alone. From our point of view, however, the principal interest of the Scott axioms is that they allow us to formulate a variant of Zermelo set theory that captures the cumulative hierarchy view of the set-theoretic universe. To obtain such a theory, it is enough to adjoin to the Scott axioms a suitable axiom of infinity asserting the existence of a limit stage.

[55]The move to a two-sorted language is just a matter of convenience; (Potter, 1990) for example contains a development of Scott's axiomatization in the standard one-sorted language of set theory.

The development of Scott's theory of ordinals is now routine. Since Scott's axiom of restriction permits one to prove that every set x is a subset of some level V, one can define $V(x)$ as the $\in$-least level V such that x is a subset of V. And then we can define the Scott ordinal of $\langle X, <\rangle$ as the set of well-ordered sets $\langle Y, <\rangle$ of $\in$-least level. In addition, it is now possible to define the usual operations of ordinal sum, multiplication, and exponentiation without appeal to the general method of transfinite recursion.[56]

To summarize, even though von Neumann's theory of ordinals and the development of transfinite recursion undoubtedly supplies us with compelling reasons to accept replacement, the beginnings of an alternative theory are available to someone who may still be reluctant to accept the axiom.

Transfinite recursion. The other important development in modern set theory that requires unrestricted replacement is the method of definition by transfinite recursion. Though the method of transfinite recursion on a well-founded relation is perfectly general, it will be enough for present purposes to consider the case of $\in$, the element-set relation. One formulation of transfinite recursion that uses first-order variables for classes states that, given a class A, perhaps a proper class, that is, a class that is not a set, if $G : A \times V \rightarrow V$ is a functional relation, then there is a unique functional relation $F : A \rightarrow V$ such that for each x in the class A:

$$F(x) = G(x, F\lceil\{y \in A : y \in x\}),$$

where $F\lceil\{y \in M : y \in x\} = \{F(y) : y \in M \wedge y \in x\}$. When the transfinite recursion theorem is reworded in the usual manner to avoid reference to classes in favor of reference to formulas, we obtain a theorem schema of Zermelo-Fraenkel set theory, and a useful theorem schema, too. The schema of transfinite recursion may be viewed as a schema for introducing defined functions; given a functional term '$F(x, y)$' implicitly defined by the recursion, the transfinite recursion schema states that we can explicitly define it by constructing a formula of the language of Zermelo-Fraenkel set theory that defines a functional relation that satisfies the recursion clause.

[56] Michael Potter has developed the theory in detail in (Potter, 1990).

Now, the interesting question is whether the general method of definition by transfinite recursion can be sustained in the absence of replacement. Two potential difficulties present themselves. One is the possible inability of different versions of Zermelo set theory to ensure existence; what guarantee do the Zermelo axioms provide that, given a definition by transfinite $\in$-recursion, there is a functional relation that satisfies the recursion? The other is their possible inability to ensure uniqueness; what guarantee do they provide that, if there is a functional relation satisfying the recursion, then there is no other such relation?

One place in which one makes essential use of replacement is in the proof that, given a transfinite definition of a functional relation on a class, there is a functional relation satisfying the recursion which is defined on all of the members of the class. In view of the results of the first part of the chapter, we can see that, quite often, the axioms of all the different versions of Zermelo set theory will not be able to ensure that, given a definition by transfinite recursion on a class, there is a functional relation defined on all the members of the class, and which satisfies the recursion clause. To illustrate this situation, consider the following definition by transfinite recursion on On, the class of all the ordinals:

$V_\alpha = \bigcup\{\mathcal{P}V_\beta : \beta < \alpha\}$,

and recall that none of the variants of Zermelo set theory we have considered have the resources to justify the existence of a functional relation which both satisfies the recursion and is defined for all the ordinals; some of them cannot prove the existence of V_ω, none of them can prove that if a limit ordinal λ exists, then so does V_λ exist. This situation is not alleviated by the addition of what we have called countable replacement. To obtain a model of Zermelo set theory plus countable replacement in which there is no functional relation satisfying the recursion, consider the structure $\langle H(\beth_{\omega_1}), \in \cap (H(\beth_{\omega_1}) \times H(\beth_{\omega_1}))\rangle$. This structure is a model of Zermelo set theory plus countable replacement which contains ω_1 as a member, but which does not contain V_{ω_1}, a set of cardinality $\beth_{\omega_1}$.

It is an intriguing question whether the axioms of Zermelo set theory have the resources to prevent more radical failures of existence, perhaps due to the existence of definable infinite descending $\in$-sequences.[57] The standard proof of the existence of a functional

[57]The fact that the infinite descending $\in$-sequences we found in models of Zermelo set theory are definable is crucial. The mere fact that there are non-well-founded models of (first-order) Zermelo-Fraenkel set is no

relation introduced by transfinite recursion on a certain class constructs the functional relation from approximations on initial ($\in$-)segments of the class satisfying the recursion, which can be shown to exist by appeal to the principle of $\in$-foundation and replacement. Now: given the fact that the principle of $\in$-induction fails in certain models of Zermelo set theory, the intriguing possibility emerges of a class none of whose members is in the domain of a partial function satisfying the recursion.[58]

What the existence of models of Zermelo set theory with definable infinite descending $\in$-sequences can be used to establish is that, given a definition by transfinite recursion, the axioms of Zermelo set theory cannot guarantee the uniqueness of a functional relation satisfying the recursion. To that purpose, we can simply use the construction $\langle V_{\omega+\omega}, \in_{new} \cap(V_{\omega+\omega} \times V_{\omega+\omega})\rangle$ of Theorem 5. For recall that this is a model in which the sets Z_0, $\{Z_0\}$, $\{\{Z_0\}\}$, ... form a (definable) infinite descending $\in_{new}$-sequence of members of $V_{\omega+\omega}$. Now: given the functional relation $y = \{x\}$, take a simple definition by transfinite recursion on the members of the infinite descending $\in_{new}$-sequence Z_0, $\{Z_0\}$, $\{\{Z_0\}\}$, ... :

$F(x) = \{F\restriction\{y \in \{Z_0, \{Z_0\}, \{\{Z_0\}\}, ...\} : y \in x\}\}$,

This recursion is supposed to define a functional relation on the class $\{Z_0, \{Z_0\}, \{\{Z_0\}\}, ...\}$ which assigns to each member of the sequence the unit, according to $\in_{new}$, of whatever it assigned to its immediate $\in_{new}$-predecessor — observe that, again according to $\in_{new}$, each member of the sequence has a unique predecessor. The trouble is that little manipulation ought to convince us that F_1 and F_2 are two different functional relations that satisfy the recursion clause:

$F_1(x) = x$. (Thus: $F_1(Z_0) = Z_0$, $F_1(\{Z_0\}) = \{Z_0\}$, $F_1(\{\{Z_0\}\}) = \{\{Z_0\}\}$, ...)

$F_2(x) = \bigcup x$. (Thus: $F_2(Z_0) = \{Z_0\}$, $F_2(\{Z_0\}) = \{\{Z_0\}\}$, ...)

Surely if F_1 is the identity function restricted to the members of the sequence Z_0, $\{Z_0\}$, $\{\{Z_0\}\}$, ..., then it satisfies the recursion; it assigns to each member of the sequence the unit set, according to $\in_{new}$, of what it assigned to its immediate $\in_{new}$-predecessor, that is,

obstacle for us to be able to use transfinite recursion, for all these models are models of all of the instances of the pinciple of $\in$-induction, too.

[58]It is not obvious, for example, that, given a definable infinite $\in$-descending sequence, one can show that there is a functional relation, rk, defined on the members of the sequence which satisfies the recursion: $rk(x = \bigcup\{rk(y) + 1 : y \in x\}$.

it assigns to it the unit set of its predecessor. But, similarly, if F_2 is the functional relation that assigns to each member of the sequence its $\in_{new}$-immediate predecessor, then it also assigns to each member of the sequence the unit set of what it assigned to its immediate $\in_{new}$-predecessor, that is, it assigns to it the unit set of the immediate $\in_{new}$-predecessor of its $\in_{new}$-predecessor, which is no other than its immediate $\in_{new}$-predecessor.

The moral seems inescapable: as they stand, common versions of Zermelo set theory are inadequate to sanction the general method of definition by transfinite recursion. Nevertheless, it is not difficult to think of possible repairs one could use to remedy this situation. What is required for the usual proof of uniqueness to ensure that, given a definition by transfinite recursion, if there is a functional relation satisfying the recursion, then it is unique is the derivability of all instances of the principle of $\in$-induction. But this could be achieved by the adoption, for example, of an axiom asserting the existence of the transitive closure of every set, or, perhaps less economically, by the adoption of countable replacement. The full exercise of replacement would still be required to ensure that, given a definition by transfinite recursion on a class, there is a functional relation which both satisfies the recursion and is defined on all the members of the class, and this should be admitted as further evidence in favor of the replacement axiom.

2.4 Conclusion

I have argued that it is a mistake to regard replacement as a mere closure postulate on the ordinal levels of the cumulative structure with few or no applications within the first $\omega + \omega$ levels of the cumulative hierarchy. In contrast with this picture, a picture of replacement has emerged as a principle of set construction which is required even to ensure that the cumulation of sets described by the axioms of set theory reaches the stage ω in the cumulative hierarchy. And we have seen that there are other important, often neglected applications of replacement are required at remarkably low levels of the cumulative hierarchy. Most of these involve instances of replacement on a countable set, and may not seem to provide us with compelling reasons to accept the full force of replacement. What they establish, however, is that it would be ill-advised to abandon the axiom in the absence of a suitable replacement.

Even after the claim that the applications of replacement are rare and exotic has been corrected, the question remains of why we should accept replacement in its unrestricted form, that is, replacement be it replacement on a countable set ot on an uncountable set, when its applications on uncountable sets still seem recondite. I have argued that the best case for replacement comes not from intuitive considerations about the concept of set, but rather from the fact that, in its absence, it is no longer possible to develop von Neumann's theory of ordinals, and, more importantly perhaps, the fact that the general method of transfinite recursion is no longer sustained. With transfinite recursion in place, one is in a position to describe and assert the content of the modern cumulative view of the set-theoretic universe, and, even though this can be indirectly be accomplished without appeal to replacement, it is plausible to suppose that no other natural addition to the axioms of Zermelo set theory achieves this. Though, in the end, these may not be conclusive reasons to in favor of replacement, they certainly show that it cannot be abandoned without a significant cost.

References

[1] Bernays, P. 1948, A System of Axiomatic Set Theory VI, *Journal of Symbolic Logic*, 13, 65-79.

[2] Boolos, G. 1989. Iteration Again. *Philosophical Topics*, 17, 5-21. Reprinted in *Logic, Logic, and Logic.*

[3] Boolos, G. 1994, The Advantages of Honest Toil over Theft. In A. George, ed., *Mathematics and Mind*, 27-44. Oxford: Oxford University Press. Reprinted in *Logic, Logic, and Logic.*

[4] Boolos, G. 1998a. Must we believe in set theory? Written for Gila Sher and Richard Tieszen, eds., *Between Logic and Intuition: Essays in Honour of Charles Parsons*, forthcoming from Cambridge University Press. Reprinted in *Logic, Logic, and Logic.*

[5] Boolos, G. 1998b. *Logic, Logic, and Logic.* Cambridge: Harvard University Press.

[6] Drake, F. 1974. *Set Theory: An Introduction to Large Cardinals*. Amsterdam: North-Holland.

[7] Felgner, U. 1971. *Models of ZF-Set Theory*. Berlin: Springer-Verlag. Lecture Notes in Mathematics 223.

[8] Fraenkel, A. 1921. Über die Zermelosche Begründung der Mengenlehre. *Jahresbericht der deutschen Mathematiker-Vereinigung*. 97-98.

[9] Fraenkel, A. 1922. Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre. *Mathematische Annalen*. 230-237.

[10] Fraenkel, A. 1925. Untersuchungen über die Grundlagen der Mengenlehren. *Matematische Zeitschrift*. 250-273.

[11] Fraenkel, A. 1926. Axiomatische Theorie der geordneten Mengen. *Journal für die reine und angewandte Mathematik (Crelle's Journal)*. 126-158.

[12] Fraenkel, A. and Yeshoua Bar-Hillel, 1958. *Foundations of Set Theory*. Amsterdam: North-Holland.

[13] Friedman, H. 1971. Higher set theory and mathematical practice. *Annals of Mathematical Logic* 2, 325-357.

[14] Hallett, M. 1984. *Cantorian Set Theory and Limitation of Size*. Oxford: Oxford University Press.

[15] Lavine, S. 1994. *Understanding the Infinite*. Cambridge: Harvard University Press.

[16] Lennes, N. J. 1922. On the Foundations of the Theory of Sets. *Bulletin of the American Mathematical Society*, 28, 300.

[17] Levy, A. 1960. Axiom schemata of strong infinity in axiomatic set theory, *Pacific Journal of Mathematics*, 10, 223-238.

[18] Levy, A. 1979. *Basic Set Theory*. Berlin: Springer-Verlag.

[19] Martin, D. 1975. Borel determinacy. *Annals of Mathematics*, 102, 363-371.

[20] Moschovakis, Y. 1994. *Notes on Set Theory*. Berlin: Springer-Verlag.

[21] Potter, M. 1990. *Sets. An Introduction*. Oxford: Oxford University Press.

[22] Rieger, L. 1957. A contribution to Gödel's Axiomatic Set Theory, *Czechoslovak Mathematical Journal*, 7, 323-357.

[23] Scott, D. 1974. Axiomatizing Set Theory. In Thomas Jech, ed. *Axiomatic Set Theory*, II, Proceedings of Symposia in Pure Mathematics.

[24] Skolem, T. 1922. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. *Matematikerkongressen i Helsingfors den 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*. References are to the translation in van Heijenoort.

[25] van Aken, 1986. Axioms for the set-theoretic hierarchy. *Journal of Symbolic Logic*, 51, 992-1004.

[26] van Heijenoort, J. 1967. *From Frege to Gödel. A Source Book in Mathematical Logic, 1879-1931*. Cambridge: Harvard University Press.

[27] von Neumann, J. 1923a. Zur Einfürung der transfiniten Zahlen. *Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae. Sectio Sci-Math.*. References are to the translation in van Heijenoort.

[28] von Neumann, J. 1923b. Letter to Ernst Zermelo, 15 August 1923. In Meschkowski, H. 1967. *Probleme des Unendlichen: Werk und Leben Georg Cantors*. Vieweg. 171-173.

[29] von Neumann, J. 1925. Eine Axiomatisierung der Mengenlehre. *Journal für die reine und angewandte Mathematik*. References are to the translation in van Heijenoort.

[30] von Neumann, J. 1928. Die Axiomatisierung der Mengenlehre. *Mathematische Zeitschrift*, 27, 669-752.

[31] Wang, H. 1970, A Survey of Skolem's Work in Logic. In Fenstad, J. E. *Selected Works in Logic*. Universitatsforlaget.

[32] Zermelo, E. 1908a. Neuer Beweis für die Möglichkeit einer Wohlordnung. *Mathematische Annalen*, 65, pp. 107-128. References are to the translation in van Heijenoort.

[33] Zermelo, E. 1908b. Untersuchungen über die Grundlagen der Mengenlehre, I. *Mathematische Annalen*, 65, 261-281. References are to the translation in van Heijenoort.

[34] Zermelo, E. 1930. Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre, *Fundamenta Mathematicae*, 16, 29-47.

Chapter 3

A No-Class Theory of Classes

In the present chapter we shall be concerned with *the* in the plural: the inhabitants of London, the sons of rich men, and so on. In other words, we shall be concerned with *classes*.

Bertrand Russell, *Introduction to Mathematical Philosophy*, p. 181.

3.1 Introduction

George Boolos (Boolos, 1984, 1985) developed a plural interpretation of second-order set theory. Boolos observed that plural quantification does not require a separate specification of the range of plural variables once the range of individual variables has been specified, and exploited this feature to argue that, if we identify second-order quantification with plural quantification, then we can both let the first-order variables of the language of second-order set theory range over all the sets there are and insist that there need not be a separate domain of second-order entities over which the second-order variables of the language range.

In this chapter, we shall make use of the apparatus of plural quantification to interpret both two- and one-sorted first-order impredicative theories of classes, much in the spirit of a suggestion of Richard Cartwright in (Cartwright, 1998). We are going to see that a plural interpretation of impredicative theories of classes has a number of advantages over other, more traditional accounts of the set-class distinction, and that it provides us with the machinery necessary to carry out recent, important developments in contemporary set theory.

Much very important and very interesting work in set theory in the last two decades has been concerned with large cardinals that have been characterized in model-theoretic terms that are not directly formalizable in the language of standard set theory, by which I mean, as usual, Zermelo-Fraenkel set theory plus the axiom of choice (ZFC). Most of these cardinals are supposed to be the least ordinal moved by a certain injective map other than the trivial identity map of the universe of all sets, V, into some transitive $\in$-model, M, of ZFC with $On \subseteq M$ (an inner model of ZFC) that preserves first-order formulas — x_1, ..., x_n satisfy the first-order formula $\phi(x_1, ..., x_n)$ if and only if their images $j(x_1)$, ..., $j(x_n)$ satisfy $\phi(v_1, ..., v_n)$ in M. The stronger the closure conditions imposed on the inner model, the stronger the corresponding large cardinal principle.

Now, one difficulty with large cardinal hypotheses asserting the existence of such cardinals is that their direct formalizability both upon the formalizability of the relation of class satisfaction for formulas of the language of ZFC, which is not quite formalizable in ZFC on account of Tarski's result on the undefinability of truth, and the existential assertion of a class that is not a set, a map j of V into some inner model of ZFC. To be sure, if the principal impediment for the formalizability of such statements consisted merely in the undefinability of the class satisfaction relation, then perhaps we could enrich the language of ZFC with a new, implicitly defined satisfaction predicate, and then proceed to formalize the necessary model-theoretic concepts in terms of that predicate. Surely we could, but a more persistent problem would remain. For the enriched language would still lack the resources necessary to assert the existence of class that is not set, such as a map j of the universe V into some inner model of ZFC.

The move to an impredicative theory of classes like Morse-Kelley set theory (MK) suggests itself. For MK provides us with the resources necessary to directly formalize the class satisfaction relation for the formulas of ZFC and to make existential assertions of the desired sort. This theory is often formulated in a two-sorted, first-order language, $\mathcal{L}$, with lowercase variables, x, y, z, ... for sets, and uppercase variables X, Y, Z, ... for classes. This language, $\mathcal{L}$, contains a two-place predicate letter, "$\in$," read: "is a member of."[1] There are two axioms that are concerned with classes:

Class Extensionality: $\forall x(x \in X \leftrightarrow x \in Y) \rightarrow X = Y$

Impredicative Comprehension: $\exists X \forall x(x \in X \leftrightarrow \phi)$

where ϕ is a formula in which the variable 'X' does not occur free. Two other axioms of the Morse-Kelley system are:

Separation: $\forall X \forall x \exists y \forall z(z \in y \leftrightarrow z \in x \wedge z \in X)$,

and

Replacement: $\forall X(X \text{ is a function} \rightarrow \forall x \exists y \forall z(z \in y \leftrightarrow \exists w(w \in x \wedge \langle w, z\rangle \in X)))$.

The rest of the axioms of MK coincide with their counterparts in ZFC.

Now, MK is both a theory suited to formalize the relation of class satisfaction for formulas of ZFC and a theory with the resources necessary to assert the existence of impredicative classes of the desired sort. Nevertheless, it should be noted that MK is a theory strictly stronger than ZFC; not only does MK formalize the relation of class satisfaction for formulas of ZFC, it even proves that there is such a class as the class of (Gödel codes of) true formulas of ZFC. Therefore, MK proves the consistency of ZFC, and thus is not a conservative extension of ZFC.

Before I comment on the difficulties involved in the interpretation of a theory like MK, let me briefly mention that another set theory that encompasses sets and classes is Gödel-Bernays set theory (GB), the theory that results from MK when the axiom schema of impredicative comprehension is weakened to:

Predicative Comprehension: $\exists X \forall x(x \in X \leftrightarrow \phi)$

[1]This is just a matter of convenience, since the theory could be formalized in a one-sorted language with variables for classes just as well.

where ϕ is a formula in which no class variables are quantified. All of the theorems of ZFC are theorems of GB, but in addition, GB is a conservative extension of ZFC: if ϕ is a formula of ZFC that is a theorem of GB, then ϕ is a theorem of ZFC as well. For if $\mathcal{M}$ is a model of ZFC that is a model of $\neg\phi$, then it is not difficult to expand it into a model of GB that is a model of $\neg\phi$ with exactly the same sets.

It is not uncommon for set theorists who make use of predicative classes to take comfort in the fact that a predicative theory like GB is a conservative extension of ZFC. For, they reason, even if classes are regarded as a valuable heuristic resource, the fact remains that predicative classes are ultimately dispensable for purposes of establishing set-theoretic facts that are formalizable in standard set theory. The problem with this view, however, is that we are often interested in facts which are not directly formalizable in standard set theory. For one example, most of the large cardinal hypotheses we have just mentioned involve the existence of elementary embeddings which are not quite definable as the range of a formula $\phi(x, y)$ of the language of ZFC.

3.2 The trouble with classes

Both Morse-Kelley (MK) and Gödel-Bernays (GB) set theories are first-order *interpreted* theories; their lowercase, or set variables of the language are supposed to range over all sets, and a formula "$x \in y$" is taken to be true (relative to an assignment of values to the variables) just in case the set assigned to the variable 'x' is a member of the set assigned to the variable 'y'. The trouble is that, in order to complete the interpretation, we need to specify both the range of the uppercase, or class variables of the language and the conditions under which a formula "$x \in Y$" is to be evaluated as true relative to an assignment of values to the variables. This is the general problem we are going to discuss in this chapter.

The first point to be noticed is that the range of the uppercase, or class variables of the language must be different from the range of the lowercase, or set variables of the language; by predicative comprehension, the sets that are not members of themselves are the members of a class, but they cannot be members of a set on pain of contradiction — we will use the term "proper class" to refer to those classes that are not sets. In other words, set theories that encompass both sets and classes must distinguish certain classes, proper classes, from sets.

This need not be a problem, provided that we interpret the lowercase, or set variables of the language to range over the members of some set — perhaps even one of the form V_κ for some suitable ordinal κ. For then, we could just interpret the class variables of the language to range over the power set of the set over which the set variables range — $V_{\kappa+1} - V_\kappa$ in the case in which the set variables of the language are taken to range over V_κ. The problem I would like us to confront in this chapter arises when we take the set variables of the language to range over all the sets there are. For it is only then that we are forced to confront the question of what are (proper) classes, if not sets.

A common picture of the set-class distinction is that, while sets are collections — where the term "collection" is used in a generic fashion to refer to entities that may do duty for sets and classes — that are formed combinatorially from their elements in the cumulative hierarchy, proper classes are collections that are "too big" to form sets. Thus, it is often remarked that the distinction between sets and proper classes is motivated by a distinction between two different conceptions of collection. According to a "combinatorial" conception of collection, collections are combinatorially formed from their elements.[2] According to a "logical" conception of collection, the characteristic mark of a collection is there is a logical collection that corresponds to each partition of the universe into two categories depending on whether objects are members of it or not. The combinatorial conception of a collection is sometimes supposed to underlie the picture of sets as formed in levels or stages of the cumulative hierarchy, and thus set theory is often viewed as the most comprehensive theory of combinatorial collections. It is, for example, in terms of this contrast that it is often explained that, while Frege and Russell were profoundly disturbed by the set-theoretic antinomies, Cantor remained unmoved by them.

[2]This is admittedly vague, as different theorists use the label "combinatorial" differently. A number theorists, most notably Penelope Maddy, use the term "combinatorial collection" to refer to sets formed in the cumulative hierarchy, but other theorists are at pains to insist that the notion of "combinatorial collection," which can presumably be traced to Cantor, is independent from the iterative conception of set, which can be argued to have played no role in the development of set theory until the late 1940s. However, even if this distinction is made, it is still plausible to suppose that the iterative conception presupposes the notion of combinatorial collection, and thus to think of iterative sets as a species of combinatorial collections.

There are important differences between the two conceptions of collection. For one, logical collections are often supposed to be more encompassing than combinatorial collections: no combinatorial collection contains all collections, but there certainly is a logical collection of all collections. And yet, according to some accounts of the distinction,[3] there is another respect in which combinatorial collections are more encompassing than logical collections. For example, there are, on these accounts, more combinatorial collections of natural numbers than there are logical collections of them. And the reason for this is that, on these accounts, a logical collection exists only as the extension of a predicate, and there are combinatorial collections — or sets — of natural numbers whose members are the extension of no predicate.

Another difference is that, if objects, logical collections would seem eligible to be members of themselves. Thus, for example, since the logical collection of all infinite collections is itself infinite, it must be a member of itself. And yet, since combinatorial collections are combinatorially formed from their elements, which must be given in advance, no combinatorial collection can contain itself as a member.

But what is perhaps the most important difference between combinatorial collections and logical collections is that the assumption that logical collections are objects gives rise to the set-theoretic antinomies. Russell's paradox arises from the assumption that there is such an object as the logical collection of non-self-membered collections. Mirimanoff's paradox arises from the assumption that there is a logical collection of well-founded collections, and the Burali-Forti paradox arises from the assumption that there is a collection of all ordinals. As a result, though the combinatorial conception of collection may be supposed to provide us with a motivation to regard set theory as the most comprehensive theory of combinatorial collections, the problem arises whether there is an intelligible and coherent account of the notion of logical collection that may help us provide an account of the set-class distinction.

The problem is not all that urgent as far as predicative theories of classes are concerned; in practice, since the axiom of comprehension of GB doesn't require the existence of classes other than those which are determined by first-order formulas with quantifiers that range

[3]For example those in (Maddy, 1990) and in (Lavine, 1994).

over sets, its classes are often treated via circumlocution in the metatheory. The situation is entirely different in the case of impredicative theories of classes. In MK, for example, there are classes that are determined by formulas with quantifiers that range over classes, and thus we must deal with the question of what exactly the items class variables range over are.

Before I comment on what is perhaps the most common solution to this problem, let me stress that the difficulty under consideration arises exclusively on the, otherwise entirely reasonable, assumption that we can quantify over all sets. For example, it is not a problem for those who adopt the point of view, which Ernst Zermelo hinted at in (Zermelo, 1930) and Charles Parsons subsequently developed in (Parsons, 1974), according to which the phrase "all sets" is hopelessly ambiguous: whenever we use it, we quantify not over all the sets there are but only over the members of a certain set — presumably a V_κ, for a sufficiently large κ. The advantage of this view is that there remains a mitigated sense in which proper classes can be reduced to sets. For there is, according to Parsons, a "higher" perspective from which what we regard as proper classes are sets which happen to fall outside the scope of our quantifiers from the original perspective. For present purposes, however, we will simply assume that we manage to quantify over all sets there are, and hence that proper classes cannot be reduced to sets even in the manner suggested by Parsons.[4]

A much more common reaction to the difficulty that concerns us is to admit that we can indeed quantify over all the sets there are, but to posit, in addition to the existence of sets, the existence of additional gigantic set-like entities, only too encompassing to be sets. Classes are then supposed to form an additional layer of the cumulative hierarchy: For example, if V is V_κ for κ an inaccessible, then some theorists conceive of the classes of MK as indistinguishable from what would be an additional layer of sets, $V_{\kappa+1} - V_\kappa$.

What is perhaps the most serious difficulty with this response is that, once we conceive of proper classes as gigantic set-like objects, only too encompassing to be sets, then we have to concede that not even the theory of classes can be the most comprehensive theory of

[4](Parsons, 1977) combines a relativistic view of the set-theoretic quantifiers with the thought that the distinction between sets and classes can be explained in terms of the different intensional principles they satisfy. Unfotunately, a discussion of Parsons' account is beyond the scope of this chapter.

collections. For just as the sets that are not members of themselves can be collected into a class, there is no reason to suppose that the classes that are not members of themselves cannot be collected into some other set-like entity, a super-class perhaps. Thus there must be a theory of super-classes some of whose variables range over these new set-like entities. But surely there is no reason to stop there: it must be possible to collect the super-classes that do not belong to themselves into a new sort of set-like entity, a hyper-class, ... And the result of course is an iteratively-generated hierarchy of class-theoretic universes at the bottom of which lie the sets recognized by set theories like ZFC. Though perfectly coherent, this is a view that strikes one as unstable, and certainly not preferable to a view of classes that is compatible with the view that regards set theory as the most comprehensive theory of set-like entities.

3.3 Plural quantification and classes

A crucial, but tacit assumption in the statement of the difficulty is that reference to classes is to be construed either as singular reference to objects of one sort or another, or else as illusory. According to this common assumption, set theorists who speak of the class of all ordinals, On, or of the class of all sets, V, for that matter, must either be taken to refer to gigantic set-like containers that are only too large to be sets or else such talk must not be taken literally. Yet, as Richard Cartwright has recently suggested in (Cartwright, 1998), another, often neglected alternative is to construe reference to classes not as singular reference to gigantic set-like entities other than sets, but as plural reference to sets. Thus, set theorists who speak of the class of all ordinals, On, may be taken to refer not to some gigantic ordinal-container, but rather to refer to the ordinals themselves in disguised notation. The aim of this chapter is to exploit the resources of plural quantification and plural reference to develop this suggestion in detail. Plural quantification, in the sense of (Boolos, 1984), provides us with the resources necessary to simulate "logical collections," and to nevertheless maintain that set theory is the most comprehensive theory of set-like entities, and not merely those lying at the bottom of a certain hierarchy of class-theoretic universes. The proposal we shall consider consists in the interpretation of each first-order formula of a theory of classes with class variables in it as a plural assertion about sets. To that purpose, we will need to resort to the apparatus of plural quantification. We shall need plural pronouns (variables), a plural predicate to link singulars with plurals: "is one of," and a plural quantifier: "There are zero or more," as explained in (Boolos, 1984).

What I would like to do now is to argue that this apparatus supplies us with the resources necessary to simulate "logical collections" and thus provide an attractive interpretation of impredicative theories of classes in which set variables are taken to range over all sets there are. To that purpose, I take my cue from recent comments on Cantor's explanation of the concept of set by Richard Cartwright. In a recent paper (Cartwright, 1998), commenting on Cantor's explanation of the concept of set, he suggests that we can make sense of Cantor's concept of an inconsistent multiplicity (collection) by taking the truth of:

(1) There are some sets that are such that no one of them is a member of itself and such that every set that is not a member of itself is one of them,

to be enough for there to be a collection of sets that are not members of themselves — enough, that is, for the truth of the sentence:

(2) There is a collection of sets such that no one of them is a member of itself and such that every set that is not a member of itself is one of them.

It is important to observe that the truth of (1) is all it is required for the use of the plural description "The non-self-membered sets" to be legitimate. Then, the thought is that perhaps we can take the use of the term "collection," in (2), to be merely a device to refer in the singular to what is perhaps more usual to refer to in the plural: the non-self-membered sets. This move will permit us to maintain that there are no set-like entities other than sets and to nevertheless admit the truth of (2). In fact, unless we take the truth of (2) not to require the existence of a set-like entity which encompasses all the non-self-membered sets, we will be forced to conclude that (2) is false.

When the term "collection" is construed as a mere device to refer in the singular to what can otherwise be referred in the plural, it would seem that, as Richard Cartwright has put it, "a collection of so-and- sos, in Cantor's sense, . . nothing over and above the so-and-sos it consists of."[5] And this is the proposal I want to consider here: to conceive of classes, be them proper or improper, as collections — or pluralities — of sets, which are nothing over and above the sets they consist of.

I should note that the idea that we construe the term "collection" as a purely singularizing device is not entirely new, as it is explicitly elaborated in (Cartwright, 1993) and it certainly echoes Bertrand Russell's "classes as many," as contrasted with what he calls "classes as one" in *The Principles of Mathematics*.[6]

An important consequence of Cartwright's proposal is that, in this purely singularizing use of the term "collection," no collection is a set.[7] This represents a departure from the use of the term "collection" by virtually all theorists as a generic term used to refer to entities that may do duty for sets and classes, but I, for one, know of no better account of the distinction between logical and combinatorial collections.

Notice that, even on this singularizing use of the term "collection," we can still make sense of the distinction between logical and combinatorial collections as a distinction that, for example, classifies some sets as a logical collection just in case they are all and only those sets that fall under one side of a certain partition of the universe. And, similarly, we can still say that some sets form a combinatorial collection just in case there is a stage of the cumulative hierarchy at which they all occur simultaneously.

The other important difference between sets and collections, in Cartwright's sense, is that, while sets are obviously members of other sets, collections cannot enter into the element-set relation, which is a relation which takes two unmistakably singular arguments: the planets of the solar system are a collection, but, as there is more than one of them, *they* cannot be *a* member of a set. And, likewise, no object can be a member of *them* — even though, of course, an object may be *one of them*. No matter: even if sets are to be

[5]Cf. (Cartwright, 1998), p. 16.

[6]Helen Cartwright writes that "collection," in this use, "serves only to singularize a plural nominal." Russell's distinction appears in (Russell, 1903), pp. 68-69. Of special interest for purposes of the present discussion is Russell's suggestion that "it is correct to infer an ultimate distinction between a class as many and a class as one, to hold that the many are only many, and are not also one" on p. 76.

[7]Modulo the concern that if x is a set, then (the collection which consists of) the sets that are identical with x would seem to be indistinguishable from the set x itself — even though not, of course, from the singleton of x.

distinguished from collections, in Cartwright's sense, it is evident that every set bears an intimate relation to the collection of its members; to use a phrase of Paul Bernays, every set *represents* a collection — in the sense that the members of the set are precisely the objects each of which is one of the collection.

From the point of view we have espoused here, only "combinatorial" collections, that is, collections that are coextensive with sets formed in the cumulative hierarchy, are represented by a set. No other collection, such as for example the collection of all sets, is represented by a set. And thus, there remains an important sense in which set theory is the most comprehensive theory of "combinatorial collections." But now, an important part of the interest of the point of view under consideration is that logical collections are no longer problematic: it is a truism that, given a division of the set-theoretic universe into two parts, there is a collection of all and only those sets on each part. Thus, what is perhaps the most salient difference between logical and combinatorial collections is that, unlike combinatorial collections, not every logical collection is *represented* by a set.[8]

The proposal I want to explore now is the suggestion that we conceive of "logical collections," or classes, as collections, in Cartwright's sense, and that, when the range of the set variables of the language of the theory of classes are taken to range over all sets, we construe the class variables of the language as plural variables that range over collections, in Richard Cartwright's sense, of sets. The principal advantage of this interpretation is that it does not require a separate specification of the range of class variables after all — hence the name "no-class theory of classes."

A plural interpretation of two-sorted Morse-Kelley set theory. The plan now is to develop a plural interpretation of a two-sorted, first-order Morse-Kelley theory of classes. This interpretation will treat the lowercase, or set variables of the language to range over the domain of all sets, but will treat the uppercase, or class variables of the language as plural variables that range over that domain. The class quantifier "$\exists X$" will be treated as a

[8]Logical collections are no longer problematic, but it may be protested that sets, objects allegedly representing collections become somewhat mysterious. Perhaps so, but this is just a version of the traditional metaphysical problem of understanding the relation a set bears to its members; what the distinction between sets and collections, in Cartwright's sense, does is to highlight an existent problem, not to generate one. Thanks are due here to Richard Cartwright.

plural quantifier in the spirit of (Boolos, 1984); that is, if ϕ^* is the result of substituting an occurrence of "$\neg x = x$" for each occurrence of "$x \in X$" in ϕ, a formula of two-sorted MK, then a formula of the form "$\exists X\phi$" of two-sorted MK is interpreted to be true if and only if either there are some sets that are such that ϕ or ϕ^*. To complete the interpretation, we must now specify the conditions under which all the different atomic formulas of the language of two-sorted MK are to be interpreted as true relative to a suitable assignment of values to the variables that occur in them.

This problem reduces to the problem of specifying the conditions under which each formula of the form "$x \in y$" and "$x \in Y$" is to be interpreted as true relative to a suitable assignment of values to the variables that occur in them. For the two place predicate '$\in$', is, in primitive notation, never flanked by another combination of set and class variables. Formulas of the form "$x \in Y$" and "$X \in Y$" are later introduced to abbreviate, respectively, the formula $\exists y(\forall z(z \in y \leftrightarrow z \in Y) \wedge y \in x)$ and the formula $\exists x(\forall z(z \in x \leftrightarrow z \in X) \wedge x \in Y)$.

The conditions under which formulas of the form "$x \in y$" and "$x \in Y$" are taken as true relative to an assignment of values to its variables are given by two clauses:

(i) A formula "$x \in y$" is true relative to an assignment of values to the variables if and only if the set assigned to the variable 'x' is a member of the set assigned to the variable 'y'.

(ii) A formula "$x \in Y$" is true relative to an assignment of values to the variables if and only if the set assigned to the variable 'x' *is one of* the sets assigned to the *plural* variable 'Y'.

Abbreviate "$\exists y(\forall z(z \in y \leftrightarrow z \in Y)$" as "*y represents* Y," that is, a set, z, *represents* a class, Y, just in case "z is the set of exactly those sets which are such that a set x is one of them just in case x is one of the Ys." Or, a bit less verbosely, "Y consists of exactly those sets that are members of z." Two immediate consequences of (i) and (ii) are:

(iii) A formula "$Y \in x$" is true relative to an assignment of values to the variables just in case the sets assigned to the *plural* variable 'Y' are *represented* by a member of the set assigned to the variable 'x',

(iv) A formula "$X \in Y$" is true relative to an assignment of values to the variables just in case the sets assigned to the *plural* variable 'X' are *represented* by one of the sets assigned to the *plural* variable 'Y'.

One technical point deserves mention: on the plural interpretation just given, the symbol "=" can no longer be taken to denote the identity relation when flanked by class variables. Identity is a relation between individuals, i.e., the relation an individual bears to itself and to no other individual. But if we construe talk of classes as plural talk of sets in disguised notation, we should not expect classes, or collections in Cartwright's sense, to enter into the identity relation. This is not much of a loss, since, on the plural interpretation just given, we can still treat the axiom of class extensionality:

Class Extensionality: $\forall x(x \in X \leftrightarrow x \in Y) \leftrightarrow X = Y$

as a definition of the symbol '=', which, in primitive notation, will disappear, except when flanked by lowercase, or set variables. Now, the fact that we use the symbol '=' indicates that we are committed to the usual requirements of reflexivity, symmetry, transitivity, and substitutivity, by which I mean, as usual, the schema: $X = Y \rightarrow \phi(X) \leftrightarrow \phi(Y)$. In primitive notation, however, all the atomic formulas in which class variables appear are of the form "$x \in Y$," and thus we can rest assured that all instances of substitutivity will be derivable from class extensionality in combination with definitions of "$X \in y$" and "$X \in Y$."

The axiom of impredicative comprehension of the Morse-Kelley system:

Impredicative Comprehension: $\exists X \forall x(x \in X \leftrightarrow \phi(x))$,

where ϕ is a formula not containing X free, will now be true just in case: either there is no set x such that $\phi(x)$ or there are some sets Xs such that a set x is one of the Xs if and only if $\phi(x)$. This should strike as a logical truth; if false, it would be the case that there is a set x such that $\phi(x)$, but there would be no sets Xs such that a set x is one of the Xs just in case $\phi(x)$, and that just could not be.

As for the other two axioms of two-sorted MK with uppercase, or class variables, the axiom of separation,

Separation: $\forall X \forall x \exists y \forall z(z \in y \leftrightarrow z \in x \wedge z \in X)$,

will, on the plural interpretation, be true just in case: given certain sets Xs, if x is a set, then there is a set y whose members are exactly those members of x which are one of the Xs. And the axiom of replacement,

Replacement: $\forall X(X \text{ is a function} \rightarrow \forall x \exists y \forall z(z \in y \leftrightarrow \exists w(w \in x \wedge \langle w, z\rangle \in X)))$.

will now read: given certain ordered pairs Xs no two of which differ in their second component but not in their first component, if x is a set, there is a set, y, of exactly those sets that appear as a second component of one of the Xs with a member of x as a first component. Thus, on the plural interpretation now developed, the strength of these axioms amounts to that of their counterparts in second-order set theory.

The remaining axioms of Morse-Kelley set theory will of course retain their customary reading. This completes the plural interpretation of the two-sorted version of the Morse-Kelley system, an interpretation that takes the lowercase, or set variables of the language to range over all sets there are, and takes the uppercase, or class variables of the language to be plural variables that range over the domain of all sets, too. We have interpreted the two-place predicate "$\in$" to denote a relation we have defined in terms of the element-set relation and the relation "is one of," that is, the relation (i) a set bears to another just in case it is a member of it, (ii) a set bears to certain sets just in case it is one of them, (iii) certain sets bear to another set just in case they are represented by some member of the set, and, finally, (iv) certain sets bear to certain other sets just in case the former sets are represented by one of the latter sets.

Effective as it is when the domain is taken to encompass all the sets there are, it is important to note that the plural interpretation of two-sorted MK is not forced upon us in circumstances in which the domain of the interpretation constitutes a set. For example, if the domain of the lowercase variables is V_κ, for a suitable κ, then it is open to us to take the uppercase variables of the language to be individual variables that range over the members of $V_{\kappa+1}$.

This set-theoretic interpretation is, in the appropriate sense, isomorphic to the plural interpretation which takes the set variables of the language to range over the members of V_κ. Both interpretations coincide in the range they assign to the set variables of the language as well as in their interpretation of the symbol '$\in$' when flanked by set variables. Now, to each $x \in V_{\kappa+1}$, a set in the range of the class variables on the set-theoretic interpretation, there corresponds a collection, in Cartwright's sense, of members of x, which is in the range of the class variables on the plural interpretation. Finally, a value of a set variable on the set-theoretic interpretation bears the interpretation of the symbol '$\in$' to a value of a class variable on the set-theoretic interpretation just in case the former is a member of the latter, that is, just in case the value of that variable on the plural interpretation is one of the sets which correspond to the latter on the plural interpretation.[9]

Furthermore, it is important to realize that the mere fact that we have provided an alternative interpretation of the formalism doesn't change the fact that the underlying logic of two-sorted MK, as most theorists conceive of it, is the standard predicate calculus. That is, the range of theorems delivered by the logic of MK will remain unchanged even after we provide a solution to the problem of specifying an interpretation of the formalism that lets us take the set variables of thc system to range over all sets. The fact remains, in particular, that the theory is satisfied in a variety of countable models of different sorts. None of these models interpret the formulas of first-order MK in terms of plurals, but some of them bear witness to important facts concerning the deductive resources of the theory, as they indicate that a variety of class-theoretic statements are not provable from the axioms of MK in the context of the standard first-order predicate calculus.

A plural interpretation of one-sorted Morse-Kelley set theory. Thus far we have developed an interpretation of a version of the Morse-Kelley system in which there are two different sorts of variables, variables for sets and variables for classes. This notation is convenient, but it disguises the fact that the underlying logic of the theory is the first-order predicate calculus. A more perspicuous notation uses only lowercase variables as class variables, and distinguishes sets from classes by the fact that the former but not the latter satisfy the formula "$\exists y\ x \in y$." A prominent example is developed in the appendix of Kelley's *General Topology*, a one-sorted, first-order theory whose variables are interpreted to range over classes. I now want to suggest that we can provide a plural interpretation of Kelley's presentation of the theory of classes.

To that purpose, it would have probably been enough to realize that each formula

[9]To make this perfectly precise would require us to code a "function" from the range of the class variables of the set-theoretic interpretation to the range of class variables of the plural interpretation as a relation that correlates a set in the range of the class variables of the set-theoretic interpretation to exactly those sets that the "function" is supposed to assign to them. The reason for this is that, presumably, no functions may take collections, in Cartwright's sense, as arguments or values.

of Kelley's one-sorted presentation of MK translates into an equivalent formula in of the standard two-sorted version of MK. To each formula of Kelley's theory of the form "$x \in y$" there corresponds a formula of the form "$X \in Y$" of two-sorted MK, which in turn abbreviates: $\exists x(\forall z(z \in x \leftrightarrow z \in X) \wedge x \in Y)$, another formula of two-sorted MK. Thus, we are in a position to make use of the plural interpretation of two-sorted MK to generate a plural interpretation of Kelley's theory.

This plural interpretation would treat all the lowercase variables of the theory as plural variables that range over the domain of all sets. The existential quantifier "$\exists x$" would similarly be treated as a plural quantifier, read: "There are zero or more," and the symbol "$\in$" will be treated as a two-place predicate letter that denotes a relation certain sets bear to certain other sets if and only if the former are represented by one of the latter sets. The result of this interpretation would take a formula of Kelley's one-sorted version of MK as true relative to an assignment of values to the plural variables of the formula if and only if its counterpart in two-sorted MK is true relative to that assignment. Thus, a formula "$x \in y$" is true relative to an assignment of values to the variables if and only if the formula "$X \in Y$" of two-sorted MK, or, equivalently, the formula "$\exists x(\forall z(z \in x \leftrightarrow z \in X) \wedge x \in Y)$" of two-sorted MK, is interpreted as true relative to that assignment to its plural variables. Otherwise put, we have that:

> A formula "$x \in y$" is true relative to an assignment of values to the variables if and only if the sets assigned to the plural variable 'x' are represented by one of the sets assigned to the plural variable 'y'.

Then, just as Kelley introduces the predicate "is a set," we can introduce a new predicate, "is represented," into the language of Kelley's theory by means of the explicit definition:

> x is *represented* if and only if $\exists y\ x \in y$.[10]

This predicate is now used not to distinguish sets from classes but to distinguish classes that are represented by sets in the domain from classes that are not so represented. After

[10]The departure from Kelley's terminology is due to the fact that, in general, classes must be distinguished from sets, as we take talk of classes as plural talk about sets in disguised notation.

this distinction is made, we are still able to rephrase all our assertions about sets as assertions about classes that are represented by a set in the domain. Thus, on the plural interpretation explored in this section, the theory of classes becomes a theory of representation.

One point to be noticed is that, on its plural interpretation, Kelley's axiom of class extensionality, $\forall z(z \in x \leftrightarrow z \in y) \leftrightarrow x = y$, would need to be strengthened in order for it to serve as a definition of the symbol "=." The reason we would like the axiom of class extensionality to serve as a definition of that symbol is again that, on the plural interpretation we have developed, "=" cannot be taken to denote the identity relation, which is a relation between individuals, but not a relation between collections, in Cartwright's sense. The trouble with treating Kelley's axiom of class extensionality as a definition, however, is that it would not guarantee the derivability of all instances of substitutivity, which is the schema: $x = y \rightarrow \phi(x) \leftrightarrow \phi(y)$. What we can do instead is to strengthen Kelley's axiom of class extensionality to:

Class Extensionality: $\forall z((z \in x \leftrightarrow z \in y) \wedge (x \in z \leftrightarrow y \in z)) \leftrightarrow x = y$,

an axiom which can in fact be used to prove each instance of the substitutivity schema.

It may now seem far from evident that the rest of the axioms of Kelley's one-sorted system are verified when interpreted in plural terms. For example, consider Kelley's axiom of impredicative comprehension:

Impredicative Comprehension: $\exists x \forall y(y \text{ is represented} \rightarrow (y \in x \leftrightarrow \phi(y)))$

where ϕ is a formula which does not contain x. I think it should be admitted that it is not obvious that, on its plural interpretation, this schema adequately expresses the content of the axiom of impredicative comprehension. And, in fact, it may even be wondered whether all instances of this schema are true. To convince oneself of the truth of the plural interpretation of impredicative comprehension, one may start by considering the plural interpretation of its counterpart in two-sorted MK:

$$\exists X \forall Y(\exists y \forall z(z \in y \leftrightarrow z \in Y) \rightarrow (\exists y \forall z(z \in y \leftrightarrow z \in Y) \wedge y \in X \longleftrightarrow \phi(Y))).$$

For consider what needs to be the case for it to be true: either there are no sets Ys which both are represented and satisfy ϕ or, otherwise, there are some sets Xs which are such that, given certain sets Ys if a set y represents the Ys, then y is one of the Xs if and only if the Ys satisfy ϕ, that is, $\phi(Y)$.

To return now to Kelley's one-sorted system, it will be useful for present purposes to treat a class abstract of the form "$\{x : x \text{ is represented } \wedge \phi(x)\}$" as a plural term referring to exactly those sets representing sets xs which are such that $\phi(x)$.[11] With this stipulation in place, we can rephrase the content of Kelley's axiom as: given a formula ϕ, there are some sets xs such that a set is one of the xs just in case it is one of $\{y : y \mathit{represented} \wedge \phi(y)\}$, or there are no sets ys such that $\phi(y)$. Or, in the singular, given a formula ϕ, there is a collection, in Cartwright's sense, $\{y : y \mathit{represented} \wedge \phi(y)\}$, which consists of exactly those sets representing a collection y such that $\phi(y)$, or there is no collection y such that $\phi(y)$. And this should strike as a logical truth, too: either there are no sets ys which are both represented and such that $\phi(y)$, or there are some sets xs such that a set is one of the xs if and only if it represents some sets ys, which are such that $\phi(y)$.[12]

As for the rest of the axioms of the theory, Kelley's axiom of subsets:

Subsets: If x is represented, then there is a y such that it is represented and for each z, if $\forall w(w \in z \rightarrow w \in x)$, then $z \in y$,[13]

is now interpreted: If the xs are represented, then there are some ys which are represented by a set and such that a set is one of the ys just in case it represents some zs such that a set is one of the zs just in case it is one of the xs. Or, in the singular, if x is a collection, in Cartwright's sense, that is represented by a set then there is a collection y that is represented by a set and such that every set representing a collection z which consists exclusively of members of x is one of the sets y consists of.

---

[11]This stipulation is quite useful in the general case, but it is not entirely adequate. For consider a class abstract of the form $\{x : x \text{ is represented} \wedge \perp\}$. This would correspond to the plural description: "The sets representing sets xs such that $\perp$," which would not seem to refer. This should not be a source of major concern, however, as class abstracts will disappear in primitive notation.

[12]A curiosity: once one realizes that all the instances of comprehension strike as logical truths, one may reflect that the second-order closure of the impredicative comprehension axiom schema strikes as a logical truth, too, which would be fine except for the little detail that, on the plural interpretation we have advanced here, it does not make sense.

[13]In set-theoretic terms, Kelley's axiom of subsets asserts the existence of a superset of the powerset of a set.

An immediate consequence of Kelley's axiom of subsets is a version of the axiom of separation of set theory: if x is represented, and z is a subcollection of x, i.e., for each $w \in z$, $w \in x$, then z is represented, too. The reason is that if x is represented, then, by the axiom of subsets, there is a collection, y that is represented and such that $z \in y$. Therefore, z is represented, too.

Abbreviate: $\{y : y \text{ is represented} \land y = x\}$ by: $\{x\}$.[14] Then, another consequence of the axiom of subsets is that if x is represented, then $\{x\}$ is represented, too. For if x is represented, then $\{y : y \text{ is represented} \land \forall z(z \in y \rightarrow z \in x)\}$ is represented and it is the case that $\forall z(z \in \{x\} \rightarrow \{y : y \text{ is represented} \land \forall z(z \in y \rightarrow z \in x)\})$.[15]

As for the rest of axioms of Kelley's presentation, they can now be formulated:

Union: If x is represented and y is represented, so is $x \cup y$

where "$x \cup y$" is taken to refer to the collection $\{z : z \in x \lor z \in y\}$, i.e., the collection consisting of exactly those sets that represent a collection which is represented by a set which is either one of the sets x consists of or one of the sets y consists of.

Substitution: If f is a function and the domain of f is represented, then the range of f is represented.

Amalgamation If x is represented, then $\{z : \exists w(z \in w \land w \in x)\}$ is represented, too.

Regularity: If $x \neq \emptyset$, then there is a class y such that $y \in x \land \forall z(z \in y \rightarrow \neg z \in x)$.

Infinity: There is a class y which is represented and such that $\emptyset \in y$ and $\forall x(x \in y \rightarrow x \cup \{x\} \in y)$.

Even Kelley's axiom of global choice is verified on the plural interpretation we have developed. Kelley's axiom of global choice:

---

[14] As defined by class extensionality, when flanked by plural variables, the symbol "=" doesn't refer to the identity relation but to the relation certain members of the domain bear to certain other members of the domain just in case a set is one of the former just in case it is one of the latter sets.

[15] One could abbreviate $\{x : x \text{ represented} \land x \neq x\}$ by: $\emptyset$, and then attempt to prove from Kelley's axiom of subsets that, if a class x is represented, then so is $\emptyset$. However, the interpretation of this term presents problems of its own, and it is better to formulate and prove this theorem in primitive notation, where no class abstracts occur. Similar remarks apply to the rest of the axioms Kelley formulates with the help of class abstracts.

Global Choice: There is a choice function f whose domain is $V - \{\emptyset\}$

reads: there are some ordered pairs fs which are such that if $x \neq \emptyset$ and the xs are represented by a set, then there are ys such that $x \in y$ — the ys are represented by one of the xs — and $\langle x, y\rangle$ is one of the fs, and, given some zs such that $x \in y$, $\langle x, z\rangle$ is one of the fs iff $y = z$. And this, again, should strike one as true.

This should convince us that we are in a position to develop the theory of classes just as Kelley does in the appendix of *General Topology*. What we have done is to argue that there is a perfectly coherent and intelligible interpretation of the theory on which the domain of (plural) quantification is taken to be the domain of all sets; all we have done is to take the variables of the theory as plural variables that range over that domain, and to interpret the predicate "$\in$" to denote the relation (a collection of) certain sets bear to (a collection of) certain other sets just in case the former sets are represented by one of the latter sets.

Alternatively, if we want to interpret the theory of classes with respect to a domain that constitutes a set, then we can take the class variables of the theory to range over subsets of the domain and we can interpret the predicate "$\in$" to denote the standard element-set relation and the predicate "is represented" to refer to all and only those subsets of the domain that are members of the domain. Indeed, since the underlying logic of Kelley's theory is still the standard first-order predicate calculus, we would expect there to be countable and other set-sized models that bear witness to the fact that a variety of statements are not theorems of Kelley's system.

Plural quantification and first-order set theories. A pleasant feature of the plural interpretation of the theory of classes is that it highlights a distinction to be made between two different roles plural variables play in the plural interpretation of impredicative theories of classes:

(i) Plural variables permit us to obtain an unexceptionable interpretation of the (unrestricted) comprehension schema.

(ii) Plural variables permit us to obtain strengthened interpretations of the axioms of separation and replacement of set theory that are immune to the expressive limitations of their first-order counterparts.

In the plural interpretation of second-order set theory, this distinction is obscured by the fact that plural variables only occur in the formulation of the axioms of separation and replacement. What I would like to do now is to suggest that this distinction is best appreciated when we consider the theories of classes that result when both the axioms of class extensionality and impredicative comprehension are adjoined to first-order theories such as ZFC; on the plural interpretation of such theories, plural variables are used merely to simulate "logical collections," but not to formulate strengthened versions of the axioms of separation and replacement.

For example, suppose the axioms of class extensionality and impredicative comprehension is adjoined to the axioms of first-order ZFC, call the two-sorted theory that results ZFC^C. The advantage of the plural interpretation we have advanced is that it is immediate to realize that ZFC^C is a conservative extension of ZFC. Both theories agree on the sentences in their common jurisdiction, and no set-theoretic fact that is not provable in ZFC is provable in ZFC^C. The reason for this is that, on the plural interpretation under consideration, the axiom of impredicative comprehension is, again, a logical truth: either there is no object x such that $\phi(x)$ or else there are some objects such that an object x is one of them just in case $\phi(x)$. Therefore, if $\mathcal{M}$ is a model of ZFC, then $\mathcal{M}$ is obviously a model of ZFC^C, too. And, conversely, if $\mathcal{M}$ is a model of ZFC^C, then $\mathcal{M}$ is trivially a model of ZFC.

It may nevertheless be useful to consider such impredicative, conservative extensions of first-order set theories just because they come equipped with the resources necessary to formalize many assertions ordinarily assumed to require impredicative classes, such as for example reflection principles, on which more later.

3.4 Substitutional quantification and predicative classes

Thus far we have advanced a plural interpretation of the Morse-Kelley system, an impredicative theory of classes. The question now immediately arises whether it is possible to specify a plural interpretation of some variant of the weaker Gödel-Bernays system that distinguishes it from Morse-Kelley set theory. Now, there is one respect in which a plural interpretation of the axioms of separation and replacement of GB would be unsatisfying. For, on their plural interpretation, the axioms of separation and replacement of GB collapse into the axioms of separation and replacement of MK, and thus, as they would be satisfied in exactly the same models, the distinction between GB and MK would become a distinction without a difference.

Fortunately, we are not necessarily required to regard the restricted quantifier of GB as plural. An alternative interpretation of the class variables that appear in the axioms of separation and replacement of GB treats them as substitutional variables that range over a specific set of formulas of first-order ZFC. As a matter of fact, substitutional interpretations of predicative theories of classes have been developed by W.V.O. Quine (Quine, 1974) and by Charles Parsons (Parsons, 1975).

To obtain a substitutional interpretation of GB, we begin with the stipulation that whenever $\phi(x)$ is a formula of the language of first-order ZFC with exactly one free variable, we rewrite $\phi(t)$ as: $t \in \{x : \phi(x)\}$. We now take X, Y, ... to be variables that range over expressions of the form $\{x : \phi(x)\}$ for $\phi(x)$ a formula of first-order ZFC, and define $X = Y$ as: $\forall x(x \in X \leftrightarrow x \in Y)$. And, finally, we introduce the substitutional quantifier $\prod X$, which ranges over formulas of the language of first-order ZFC with one free variable. Now, on the substitutional interpretation a sentence of the form $\prod X \psi(X)$ is true if and only if all sentences gotten from the schema $\psi(X)$ when "X" is substituted by an expression of the form $\{x : \phi(x)\}$, for $\phi(x)$ a formula of first-order ZFC with exactly one free variable, are true.

Thus, for example, the axiom of separation of GB can now be written:

$$\prod X \forall x \exists y \forall z (z \in y \leftrightarrow z \in x \land z \in X),$$

which will be true just in case all sentences obtained by substituting a term of the form $\{x : \phi(x)\}$, where $\phi(x)$ is a formula of first-order ZFC with exactly one free variable, for "X" in the formula $\forall x \exists y \forall z (z \in y \leftrightarrow z \in x \land z \in X)$ are true.

It is evident that the strength of the substitutional versions of the axioms of separation and replacement of GB approximates that of their schematic counterparts in first-order ZFC, but, unfortunately, they don't quite match it. For, given our account of substitutional quantification, the only formulas of ZFC in the substitution class are formulas containing exactly one free variable, but not formulas containing additional free variables or parameters. Charles Parsons (Parsons, 1971) devised a generalization of substitutional quantification that can be used to accommodate these formulas too. The thought is that, given the usual definition of satisfaction for the formulas of first-order ZFC, we can stipulate that a sequence s satisfies a formula of the form $\prod X\psi(X)$ if and only if s satisfies every formula obtained by substituting a term of the form $\{x : \phi(x, y_1, ..., y_n)\}$, where $\phi(x, y_1, ..., y_n)$ is a formula of first-order ZFC containing variables y_1, ..., y_n free, for "X" in ψ. This generalization of substitutional quantification, which is discussed by Charles Parsons both in (Parsons, 1971) and in (Parsons, 1974), gives us precisely what we wanted: an interpretation of GB that sits well with the view that there are no set-like entities other than sets.

It may be of interest to mention that, since, on the substitutional interpretation, GB is satisfied in every model of first-order ZFC, the contrast between the plural and the substitutional interpretations of GB would seem to reflect the contrast between first- and second-order ZFC.

Unfortunately, the substitutional interpretation of the class quantifiers is not available for the purpose of interpreting impredicative theories of classes, such as, for example, the Morse-Kelley theory. First off, it should be obvious that a substitutional interpretation of the class quantifiers on which a class variable "X" is supposed to range over expressions of the form $\{x : \phi(x)\}$, where $\phi(x)$ is a formula of ZFC, will inevitably fail to verify all the instances of the impredicative comprehension schema of the Morse-Kelley theory. The reason is quite simple. For notice, for example, that the class of (Gödel codes of) true formulas of ZFC can be defined as the extension of a formula of the form "$\sum X\ S(X, x)$," where $S(X, x)$ contains no class variables other than "X." But then, since:

$$\sum X \forall x (x \in X \leftrightarrow \sum X\ S(X, x))$$

is an instance of impredicative comprehension, and, by Tarski's theorem on the undefinability of truth, there is no class abstract of the form $\{x : \phi(x)\}$, where $\phi(x)$ is a formula of the language of ZFC, that is coextensive with $\sum X\ S(X, x)$, we have that the substitutional interpretation just proposed cannot verify all the axioms of Morse-Kelley set theory.

It will perhaps be suggested that we need only refine the interpretation by allowing substitutions of a class variable "X" by expressions of the form $\{x : \phi(x)\}$ in which $\phi(x)$ is not necessarily a formula of ZFC, but perhaps a formula that contains itself the substitutional quantifier $\prod X$. But now, think what might happen. A sentence of the form $\prod X\phi(X)$ is supposed to derive its truth conditions from the truth conditions of its instances, but the substitution instances of that sentence need not be simpler than the sentence itself. And thus its evaluation may lead us into circularities.[16]

To illustrate this point, let us expand the language of two-sorted MK to contain class abstracts of the form $\{x : \phi(x)\}$, where $\phi(x)$ is a formula of the language of MK, and take the class quantifier of MK to be substitutional. Now consider a sentence like:

$$\textstyle\prod X(\exists x(x \in X) \to \exists x(x \in X \wedge \forall y(y \in x \to \neg y \in X))).$$

Whether or not this sentence is true depends on whether all the substitution instances of "$\exists x(x \in X) \to \exists x(x \in X \wedge \forall y(y \in x \to \neg y \in X))$" are true. All the substitution instances of this sentence are unproblematically true except perhaps for the ones that result when "X" is substituted for a class abstract that contains the substitutional quantifier $\prod X$. For let "A(x)" abbreviate: "$\neg \prod X(x \in X \to \exists x(x \in X \to \forall y(y \in x \to \neg y \in X)))$," and consider:

$$\exists x(x \in \{x : \mathrm{A}(x)\}) \to \exists x(x \in \{x : \mathrm{A}(x)\} \wedge \forall y(y \in x \to \neg y \in \{y : \mathrm{A}(y)\})).$$

If we decide that "$\prod X(\exists x(x \in X) \to \exists x(x \in X \wedge \forall y(y \in x \to \neg y \in X)))$" is true, then the antecedent of this conditional will be false, and thus the conditional will be true.

But if we decide that "$\prod X(\exists x(x \in X) \to \exists x(x \in X \wedge \forall y(y \in x \to \neg y \in X)))$" is false, then the antecedent of this conditional will be true and its consequent false. To clearly see this, observe that if $a \in \{x : \neg \prod X(x \in X \to \exists x(x \in X \wedge \forall y(y \in x \to \neg y \in X)))$ and $\{x : \phi(x)\}$ is a class abstract that can be substituted for "X" to obtain a true substitution instance of "$\neg \prod X(a \in X \to \exists x(x \in X \wedge \forall y(y \in x \to \neg y \in X)))$," then, if $b \in a$, the class

[16]This problem is by no means specific to the language of classes, but rather it is a perfectly general difficulty: whatever one's language, if one expects to make unambiguous use of substitutional quantification, one should not allow for the occurrence of substitutional quantifiers within the substituted terms.

abstract $\{x : \phi(x) \vee x = b\}$ can be substituted by "X" to obtain a true substitution instance of $\neg \prod X(b \in X \rightarrow \exists x(x \in X \wedge \forall y(y \in x \rightarrow \neg y \in X)))$. We conclude that the conditional is false, and hence that "$\prod X(\exists x(x \in X) \rightarrow \exists x(x \in X \wedge \forall y(y \in x \rightarrow \neg y \in X)))$" is itself false.

Not surprisingly, then, though perfectly suited for the interpretation of predicative theories of classes such as GB, the substitutional interpretation of the class quantifiers is no alternative to the plural interpretation of impredicative theories of classes such as MK that we have developed here.

3.5 Proper classes and set-theoretical practice

We have seen that the plural interpretation of impredicative theories of classes presents us with a principled explanation of the set-class distinction that is compatible with the universality of set theory as the most comprehensive theory oι combinatorial collections, when these are taken to be objects. But not only are plural interpretations of theories of classes satisfactory as a principled explanation of the set-class distinction, they provide us with the machinery necessary to carry out important developments in set theory. This section will focus on just a few examples from contemporary set theory.

I remarked in the introduction that contemporary set theory is permeated with classes; a variety of large cardinal principles are formulated in model-theoretic terms that are not directly formalizable in the language of ZFC. What is perhaps the most prominent example of a model-theoretic concept used in the formulation of large cardinal principles that requires the machinery of proper classes for its formulation is that of an elementary embedding for class structures. An elementary embedding of a class structure, $\mathcal{M}_0$, into another, $\mathcal{M}_1$, is an injective map, j, from the domain of $\mathcal{M}_0$, a proper class, into the domain of $\mathcal{M}_1$, another proper class, such that:

$$\mathcal{M}_0 \models \phi(x_1, ..., x_n) \textit{ iff } \mathcal{M}_1 \models \phi(j(x_1, ..., x_2))$$

whenever $\phi(v_1, ..., v_n)$ is a formula of ZFC. j is called non-trivial if it is not the identity map, in which case $j(\delta) > \delta$ for some ordinal δ, its *critical point*. One reason for the unformalizability of the concept of elementary embedding for class structures in the language of ZFC is that, by the Gödel-Tarski undefinability of truth argument, we have that neither the satisfaction relation for formulas of ZFC, $Sat(\ulcorner\phi\urcorner, \langle x_0, ..., x_n\rangle)$ *iff* $V \models \phi$, nor the satisfaction relation for formulas of ZFC for proper classes, $Sat(\ulcorner\phi\urcorner, \langle x_0, ..., x_n, M\rangle)$, can be defined within ZFC. This is also the source of the strict unformalizability of assertions such as "$\mathcal{M}$ is an inner model of ZFC," when $\mathcal{M}$ is a class structure, but it is not the principal reason one can use to motivate the move to an impredicative theory of classes. For, after all, a more economical solution to this problem would be to supplement the language of ZFC with the addition of a satisfaction predicate implicitly defined by the usual recursive definition. What is the principal impediment for the formalizability of large cardinal hypotheses asserting the existence of an elementary embedding, j, of the universe into some inner model of ZFC arises from the inability of ZFC to assert the existence of a proper class, the map j, which need not be definable as the range of some formula, $\phi(x, y)$ of ZFC.

Despite both the strict unformalizability of the model-theoretic concept of elementary embedding for class structures and the inability of ZFC to formalize existential assertions of a proper class, set theorists have studied an entire hierarchy of large cardinal principles that assert the existence of an elementary embedding of the universe of all sets into inner models that satisfy different closure conditions. The first large cardinal principle of this sort states the existence of a non-trivial elementary embedding of the universe of all sets, V, into some inner model, a principle that is equivalent to the existence of a measurable cardinal. Other large cardinal principles are the result of the imposition of additional conditions on the inner model M above. The stronger the closure conditions imposed on M, the stronger the corresponding large cardinal principle. For example, a cardinal κ is γ-strong iff there is a non-trivial elementary embedding $j : V \to M$, κ is the least ordinal moved by j and $\gamma < j(\kappa)$, and $V_{\kappa+\gamma} \subseteq M$. A cardinal κ is strong iff it is strong for every γ. But κ is superstrong if there is a non-trivial elementary embedding $j : V \to M$ with critical point κ such that $V_{j(\kappa)} \subseteq M$.

These cardinals are only the first stages of a hierarchy that has been intensely studied in recent times. In the later stages of the hierarchy are the compact, supercompact, and huge cardinals. For example, a cardinal is γ-supercompact iff it is the least ordinal moved by a non-trivial elementary embedding $j : V \to M$ such that $j(\kappa) > \gamma$ and M is closed under γ-sequences; a cardinal κ is supercompact iff it is γ-supercompact for all $\gamma < \kappa$. There is an upper limit on the hierarchy, whose formulation is due to William Reinhardt: κ is a Reinhardt cardinal iff κ is the least ordinal moved by a non-trivial elementary embedding $j : V \to V$. In 1971, Kenneth Kunen proved that there is no non-trivial elementary embedding of the universe of all sets into itself, and hence that there are no such cardinals as Reinhardt cardinals. Ever since Kunen proved his result, an important number of theorists have concentrated their efforts in the examination of cardinals that seem as large as possible but are nevertheless immune to Kunen's argument.

It is not uncommon for set theorists to take comfort in the fact that most of the large cardinal principles I have mentioned admit of alternative characterizations that are formalizable in ZFC. For example, the existence of a non-trivial elementary embedding, j, of the universe into some inner model is equivalent to the existence of a witnessing ultrafilter on κ, the least ordinal moved by j, which is of course formalizable in ZFC. Similarly, there are alternative characterizations of supercompactness in terms of ultrapowers that are likewise formalizable in ZFC. This reaction is not entirely satisfactory. Except perhaps for the principle that asserts the existence of a measurable cardinal, the intelligibility of the large cardinal principles formulated in terms of the existence of elementary embeddings derives from their model-theoretic characterization, and not from the set-theoretic characterizations set theorists obtain a posteriori, after the theory of these principles is thoroughly developed. Not only does the use of class-theoretic characterizations taps a vital source of intuitions about these large cardinals, quite often, it is required to render arguments intelligible. This fact alone makes it desirable to be able to develop the theory of large cardinals in a setting in which it is possible to formalize the class-theoretic characterizations we have mentioned.

Furthermore, it would be hopeless to suppose that all the cardinal principles set theorists entertain will at some point reveal themselves to be equivalent to assertions that are formalizable in ZFC. For one example, Vopěnka's principle, i.e., the principle that, given a proper class of structures for a language, there is one that is elementary embeddable into another, is not formalizable in ZFC. I think we should admit, as a consequence, that ZFC is definitely not the best setting for the development of the study of the large cardinal principles mentioned thus far.

In practice, with perhaps a few exceptions, set theorists tend to develop the study of large cardinals within an informal theory of classes, and, then, after theory is developed, to consider possible, more roundabout formalizations of all these developments in the context of ZFC. The main purpose of the plural interpretation we have developed is to provide the resources necessary to sustain a well-entrenched practice, the use of the theory of classes in the study of large cardinals. One problem with the use of a theory of classes is that what is perhaps the most common interpretation of the theory involves commitment to set-like entities that are not sets. And this commitment is in tension with the universality of set theory as the most comprehensive theory of combinatorial collections, as objects, and makes the policy of disallowing proper classes to be members of other classes artificial. The plural interpretation we have developed gives us what we want: an interpretation of Morse-Kelley set theory that is compatible with the assumption that sets are all the combinatorial collections, as objects, there are, and which nevertheless accounts for the fact that proper classes cannot belong to other classes.

This interpretation makes sense of all the different model-theoretic concepts we want to formalize as well as of all the different large cardinal hypotheses we have considered thus far. Thus, for example, an inner model M is, on the plural interpretation developed here, given by certain ordered pairs — the interpretation of the two-place predicate letter "$\in$" — satisfying certain conditions: every ordinal occurs as a component of one of them, and all the axioms of ZFC are satisfied when interpreted with respect to the sets that occur as a component of one of the pairs, provided that a formula "$x \in y$" is interpreted as true relative to an assignment iff the set assigned to "x" appears as the first component of an ordered pair whose second component is the set assigned to "y." Similarly, the existence of an elementary embedding j of the universe of all sets into some inner model amounts, on the plural interpretation presented here, to the existence of certain ordered pairs satisfying the usual conditions.

Another context in which classes take center stage involves the use of certain reflection principles. Most set theorists who have written on axioms of infinity and the structure of the set-theoretic universe regard it as plausible to suppose that V, the universe of all (pure) sets, is structurally undefinable, and thus that no structural property of V fails to be reflected lower down in some level V_κ of the cumulative hierarchy of sets. For example, since On is undoubtedly strongly inaccessible — if λ is an ordinal, then 2^λ is an ordinal, and the limit of arbitrary sequences of ordinals of length $< On$ is certainly an ordinal, therefore there must be a level of the cumulative hierarchy, V_κ, with κ strongly inaccessible. But then, since On is strongly inaccessible but $> \kappa$ for a strong inaccessible κ, there must be another inaccessible, etc. Reflection is the main heuristic advanced for various large cardinals, but it should be evident that it is not formalizable in the language of ZFC.[17] Once again, the interest of the plural interpretation of the theory of classes is that it permits us to take such reflection arguments seriously, for, even if we admit that there are no such objects as proper classes, the plural number provides us with an interpretation of the language of classes on which it is perfectly coherent and intelligible to talk about structural properties of On or V. Thus, the assertion that On is strongly inaccessible is just the assertion that the ordinals enjoy certain closure properties: if λ is an ordinal, 2^λ is an ordinal, too, and the limit of sequences of ordinals that are isomorphic to some ordinal is itself an ordinal.

There is a different, but complementary approach "from below" to the study of the structure of the cumulative hierarchy in which proper classes are supposed to play an important role, too. This is an approach that is motivated by the view that set theory can be taken as a formal extension of known facts in finite set theory into the transfinite. Harvey Friedman has partially developed this approach in "Transfer Principles in Set Theory."[18] A transfer principle is an assertion to the effect that a certain fact of finite set theory can be generalized into the transfinite. In particular, Friedman has explored connections between functions on ω and functions on the entire class of ordinals, On to isolate plausible transfer principles of the form:

> If for all suitable functions f_1, ..., f_p from $N^k \to N$, $A(f_1, ..., f_p)$, then for all suitable functions f_1, ..., f_p from $On^k \to On$, $A(f_1, ..., f_p)$,

for appropriate existential sentences $A(f_1, ..., f_p)$.

[17]To the limited extent to which reflection is formalized in first-order ZFC as the principle:
$\forall\alpha\exists\beta > \alpha\forall x_1, ..., \forall x_n \in V_\beta(\phi(x_1, ..., x_n) \leftrightarrow \phi_{V_\beta}(x_1, ..., x_n))$
it is a theorem of ZFC.

[18]Cf. (Friedman, 1997).

The first point to be noticed is that no instance of this schema can be formalized in the language of ZFC, as its consequent states that a certain fact about arbitrary functions on On obtains. The interest of these transfer principles is that they provide us with information concerning both the width and the depth of the cumulative hierarchy, as some of them are equivalent with large cardinal principles that contradict the axiom of constructibility, i.e., $V = L$. In particular, Friedman has proved that different transfer principles are equivalent to different hypotheses concerning structural properties of On, such as the hypotheses that On is weakly compact, or ineffable, or Ramsey.

For present purposes, however, the interest of this development is that, again, it would seem to require for its intelligibility an adequate interpretation of the theory of classes, as none of the class-theoretic hypothesis concerning On is provably equivalent to a purely set-theoretic sentence. And I think that the plural interpretation of the Morse-Kelley theory provides us with what we want.

3.6 Conclusion

To summarize, then, we have argued that plurals provide us with interpretations of several theories that encompass sets and classes, which are not subject to the difficulties faced by rival accounts of the set-class distinction. In particular, I have suggested that the identification of classes with collections, in Richard Cartwright's sense, has a number of advantages over other, more traditional accounts of the distinction. One important advantage of the present account is that it draws a principled distinction between sets and classes, and not one based merely on size or location in the cumulative hierarchy. Plural quantification enables us to simulate "logical collections" as collections that divide the universe into two categories depending on whether objects are one of them or not, and it does so in a way in which the connection between sets and logical collections, or classes, is perfectly clear: set theory is concerned with *all* the sets — indeed all the set-like entities there are, and classes are simply collections of those entities.

Another advantage of the proposal is that it provides a rationale for the policy of disallowing proper classes to bear the element-class relation to other classes. That classes, in general, cannot bear the element-set relation to other classes is clear, since the element-set relation takes only singular arguments. But it should be equally clear that the first argument of the element-class relation, "is one of," is unmistakably singular, and hence that it can never be occupied by a class. We conclude that the requirement that classes don't bear the element-class relation to other classes is a principled one, when one conceives of classes as collections, in Richard Cartwright's sense. I should note, however, that this observation leaves open the possibility that there be a relation of the form "are some of them" whose last argument is doubly plural: it takes collections of collections of sets. Then, as Russell once put it, "a class of classes [would be] many many's; its constituents [would] each be many, and [could] not therefore in any sense be single constituents." (Russell, 1903, p. 516). Whether this is so much as an intelligible view is, however, a question that I shall not address here.

At all events, what is perhaps the most important advantage of the approach we have advocated is that it permits us to embrace the theory of classes without commitment to set-like entities other than sets, and hence without threatening the universality of set theory. As a result, on the plural interpretation we have developed here, Morse-Kelley set theory provides us with the machinery necessary to carry out important developments in contemporary set theory.

References

[1] Bernays, P. 1937-1954, A System of Axiomatic Set Theory. *Journal of Symbolic Logic*, 1937: 2, 66-77; 1941: 6, 1-17; 1942: 7, 65-89, 133-145; 1943: 8, 89-106; 1948: 13, 65-79; 1954: 19, 81-96.

[2] Boolos, G. 1984, To Be is To Be the Value of a Variable (or Some Values of Some Variables). *Journal of Philosophy*, 81, 430-450. Reprinted in *Logic, Logic, and Logic.*

[3] Boolos, G. 1985, Nominalist Platonism. *The Philosophical Review*, 94, 327-344. Reprinted in *Logic, Logic, and Logic.*

[4] Boolos, G. 1998. *Logic, Logic and Logic.* Cambridge: Harvard University Press.

[5] Cartwright, H. 1993. On Plural Reference and Elementary Set Theory. *Synthese*, 96, 201-254.

[6] Cartwright, R. 1998. A Question about Sets. Unpublished manuscript.

[7] Friedman, H. 1997. Transfer Principles in Set Theory. Unpublished draft available at: http://www.math.ohio-state.edu/foundations/manuscripts.html.

[8] Kelley, J. L. 1955. *General Topology*. London: Van Nostrand.

[9] Lavine, S. 1994. *Understanding the Infinite*. Cambridge: Harvard University Press.

[10] Maddy, P. 1990. *Realism in Mathematics*. Oxford: Oxford University Press.

[11] Parsons, C. 1971. A Plea for Substitutional Quantification. *Journal of Philosophy*. 68, 231-137. Reprinted in *Mathematics in Philosophy: Selected Essays*.

[12] Parsons, C. 1974. Sets and Classes. *Nous*, 8, 1-12. Reprinted in *Mathematics in Philosophy: Selected Es*

[13] Parsons, C. 1977. What is the Iterative Conception of Set? In R.E. Butts and Jaako Hintikka, eds. *Logic, Foundations of Mathematics, and Computability Theory*. Dordrecht: Reidel. Reprinted in *Mathematics in Philosophy: Selected Essays*.

[14] Parsons, C. 1983. *Mathematics in Philosophy: Selected Essays*. Ithaca: Cornell University Press.

[15] Quine, W.V.O. 1974. *The Roots of Reference*. La Salle: Open Court.

[16] Quine, W.V.O. 1969. *Set theory and its Logic*. Cambridge: Harvard University Press.

[17] Russell, B. 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press.

[18] Russell, B. 1919. *Introduction to Mathematical Philosophy*. London: Allen & Unwin.

[19] Zermelo, E. 1930. "Über Grenzzahlen und Mengenbereiche." *Fundamenta Mathematicae*, 16, 29-47.