The Search for New Axioms

by

Peter Koellner

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
[June 2003]
May 2003

Author .....
Department of Linguistics and Philosophy
May 16, 2003

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

Accepted by .....
Vann McGee (Chairman, Department Committee on Graduate Students), Michael Glanzberg (Assistant Professor of Philosophy, MIT), Akihiro Kanamori (Professor of Mathematics, Boston University)

The Search for New Axioms

by

Peter Koellner

Submitted to the Department of Linguistics and Philosophy
on May 16, 2003, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Philosophy

Abstract The independence results in set theory invite the search for new and justified axioms. In Chapter 1 I set the stage by examining three approaches to justifying the axioms of standard set theory (stage theory, Gödel's approach, and reflection principles) and argue that the approach via reflection principles is the most successful. In Chapter 2 I analyse the limitations of ZF and use this analysis to set up a mathematically precise *minimal hurdle* which any set of new axioms must overcome if it is to effect a significant reduction in incompleteness. In Chapter 3 I examine the standard method of justifying new axioms—reflection principles—and prove a result which shows that no reflection principle (known to be consistent via large cardinals) can overcome the minimal hurdle and yield a significant reduction in incompleteness. In Chapter 4 I introduce a new approach to justifying new axioms—extension principles—and show that such principles can overcome the minimal hurdle and much more, in particular, such principles imply PD and that the theory of second-order arithmetic cannot be altered by set size forcing. I show that in a sense (which I make precise) these principles are inevitable. In Chapter 5 I close with a brief discussion of *meta-mathematical* justifications stemming from the work of Woodin. These touch on the continuum hypothesis and other questions which are beyond the reach of standard large cardinals.

Thesis Supervisor: Vann McGee
Title: Professor

Acknowledgments

It is a pleasure to look back and think of everyone who encouraged and helped me along the path which led to this thesis. I would like to thank my father for asking me my first philosophical question; my mother and my sister for their confidence in me; my grandparents for all the jokes about how I was going to be a student for the rest of my life; John Slater and William Demopoulos for guidance; Alasdair Urquhart for being my model of a logician-philosopher; Sy Friedman for sparking my interest in set theory; SSHRC for financially supporting two of my years in Berkeley; John Steel for teaching me inner model theory; Hugh Woodin for sharing his insights into set theory and devoting so much time to my education; Vann McGee, Michael Glanzberg, and Akihiro Kanamori for being an excellent dissertation committee; Bill Tait for fruitful correspondence; Richard Ketchersid for many wonderful conversations; and Iris Einheuser for advice, support and happiness, from the time of our first class in logic to the present.

Contents

1 The Standard System **13**

1.1 The Traditional Approach 13

1.1.1 Primitive Notions 13

1.1.2 The Iterative Hierarchy 15

1.1.3 Nature 16

1.1.4 Extent 17

1.1.5 Two Conceptions: Actualist and Potentialist 17

1.1.6 Stage Theory 20

1.2 Gödel's Approach 21

1.2.1 The Account of 1933 21

1.2.2 The Account of 1951 24

1.2.3 Choice 28

1.3 Reflection Principles 29

1.3.1 Motivation 29

1.3.2 Formulation 30

1.3.3 First Order 31

1.3.4 Second Order 32

1.3.5 Conclusion 34

2 Independence **35**

2.1 Limitations 35

2.1.1 The Incompleteness Theorems 35

2.1.2 Benign Limitations 36

2.1.3 Serious Limitations 37
2.1.4 Example 37
2.2 The Minimal Hurdle 38
2.2.1 Degrees of Interpretability 38
2.2.2 The Scope of ZF 39
2.2.3 Absoluteness 40
2.2.4 The Minimal Hurdle 41

3 Reflection Principles **43**
3.1 The Program for Reflection Principles 44
3.1.1 Indescribables 44
3.1.2 The Program 45
3.1.3 Limitations for the Actualist 46
3.1.4 The Potentialist View 48
3.2 Higher-Order Parameters 50
3.2.1 Relativisation 50
3.2.2 Inconsistency 50
3.2.3 Two Fixes 51
3.3 Tait 51
3.3.1 Γ_n-reflection 52
3.3.2 Ineffables 53
3.3.3 Connection 54
3.3.4 Consistency 54
3.4 Limitations 55
3.4.1 The First Fix 55
3.4.2 The Second Fix 60
3.4.3 Conclusion 61
3.5 Appendix 61

4 Extension Principles **69**
4.1 Reinhardt 69

4.1.1 Introduction 69
4.1.2 Ackermann Set Theory 70
4.1.3 The Principle of Sharp Delimitation 71
4.1.4 Comparing the Candidates 73
4.1.5 Criticism 75
4.2 The Principle EP 75
4.2.1 Tracking Definable Subsets 75
4.2.2 The Principle 76
4.3 The Strength of EP 78
4.3.1 Freezing 78
4.3.2 Sharps 79
4.3.3 Strength 81
4.3.4 Inevitability 83
4.4 The Principle GEP 84
4.4.1 The Generalised Extension Principle 84
4.4.2 Projective Determinacy 85
4.4.3 Freezing Second Order Arithmetic 87

5 Beyond Large Cardinals **89**
5.1 Freezing 89
5.1.1 The Theory of $L(\mathbb{R})$ 89
5.1.2 Overlapping Consensus 90
5.1.3 Limitations 91
5.2 Ω-Logic and the Continuum Hypothesis 91
5.2.1 Strong Logics 91
5.2.2 Gödel's Program 93
5.2.3 The Continuum Hypothesis 94
5.3 The structure Theory of $L(V_{\lambda+1})$ 96
5.3.1 Analogy 96
5.3.2 Conclusion 97

Introduction

What are the mathematical truths? There was a time when the answer to this question was thought to be straightforward: the truths of mathematics are those statements that are derivable from a sufficiently rich set of axioms. A candidate for such a set of axioms is ZF. These axioms are extremely powerful; they are strong enough to develop all of current mathematics and constitute what is regarded as the *standard system of mathematics.* Can *all* of the mathematical truths be derived from the standard system of mathematics? Surprisingly they cannot! This remarkable result is due to Gödel, Cohen, and others. They showed that there are central problems of mathematics which are not merely *tough* to figure out—a phenomenon common to all subjects—but which are *in principle impossible* to figure out on the basis of the standard axioms of mathematics. One must therefore seek out and justify new axioms if one is to come closer to capturing mathematical truth. This is the task I undertake in this thesis.

I will begin in Chapter 1 by examining three approaches to justifying the standard system ZF: the traditional approach, Gödel's approach, and the approach in terms of reflection principles. After pointing out that the traditional approach is unsuccessful I will argue (i) that Gödel's approach yields a justification of ZF but only at the cost of assuming AC and (ii) that the approach in terms of reflection principles yields a justification of ZF (and more) without assuming AC. This naturally leads to the question of whether reflection principles can justify extensions of ZF that are strong enough to yield a "significant" reduction in incompleteness. Chapter 2 is devoted to making precise the notion of a "significant" reduction in incompleteness and to setting up a mathematically precise Minimal Hurdle which any set of new axioms

must overcome if it is to effect a significant reduction in incompleteness. This puts us in a position where we can prove results of the form: axioms of type A can (or cannot) yield a significant reduction in incompleteness. The first result of this form occurs in Chapter 3 where I show that no *reflection principle* known to be consistent can overcome the Minimal Hurdle. This limitative result encourages us to look for a new method of justifying axioms. In Chapter 4 I introduce a new method for justifying axioms and use it to derive two *extension principles*: EP and GEP. I show that EP overcomes the Minimal Hurdle and that GEP yields a much greater reduction in incompleteness—in particular, it implies PD and freezes the theory of second order arithmetic. In the final chapter I chart out the limitations of this approach and begin an examination of a new—*metatheoretic*—method of justification.

Chapter 1

The Standard System

Our starting point is the standard axioms of set theory and the various justifications for these axioms. §1.1 begins with a discussion of the fundamental notions of set theory and concludes with a criticism of the standard justification of ZF. Along the way I draw a distinction between two conceptions of the universe of sets—*actualism* and *potentialism*, a distinction that will play a central role in subsequent chapters. §1.2 argues that an approach due to Gödel (in unpublished work) yields a justification of ZF but only at the cost of assuming AC. In §1.3 I argue that reflection principles give a uniform and streamlined justification of ZF (and more) without presupposing AC. These principles will be pursued further in Chapter 3 where I chart out the extent to which one can found extensions of ZF on their basis.

1.1 The Traditional Approach

1.1.1 Primitive Notions. There are two primitive notions in the foundations of set theory. The first primitive notion is the operation of *set formation*, which I shall abbreviate 'set of' or, more explicitly, 'set of x's' where "the variable x ranges over some given kind of objects" (Gödel (1964)). The 'set of' operation takes a given kind K of objects and forms a single object S, called a 'set', which has as members exactly those objects of kind K. Another way to state this is in terms of collections of objects considered as *many*: the 'set of' operation takes a collection C of objects considered

as many and produces a unity S. The advantage of phrasing matters in this second way is that it makes it clear that the objects of the given kind are already available, thus distinguishing it from the naive principle of comprehension that leads to Russell's paradox. The disadvantage is that when one speaks of a collection considered as many (what I shall refer to as a *plurality*) it is tempting to think that one is talking about an object. I will use the terminology of pluralities but I want to stress two things: first, that a plurality is not an object, it is an array of objects, and, second, that all talk of pluralities can be rephrased in terms of given kinds.

The relation 'a is a member of S' is abbreviated '$a \in S$'. For example, consider the natural numbers 0, 1, 2, 3, ... Applying the operation 'set of' to this given kind or plurality we obtain the set $\{0, 1, 2, 3, \ldots\}$, an object which has as members all and only the natural numbers. This object is denoted ω.

The second primitive notion is the operation of *powerset*. This is the operation which, given a set X, produces the set $\mathscr{P}(X)$ of *all arbitrary* subsets of X. Here an *arbitrary* subset is to be contrasted with a *definable* subset. As examples of definable subsets of ω consider the set of even numbers and the set of odd numbers. Since there are only countably many definitions there are only countably many definable subsets of ω. In contrast there are uncountably many arbitrary subsets of ω.

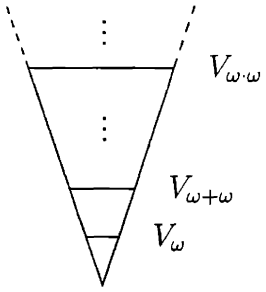
Three comments are in order. (1) In formulating the operation of set formation I said that it forms a set of any "given kind" of objects. It follows that the collection of "all sets" is not a "given kind" for if it were then we should be able for form the set of all sets which is impossible. For those who think that "the sets" constitute a given kind this simple fact will be taken to demonstrate that the operation of set formation is *not* a primitive of set theory. Below in §1.1.5 I will argue that the inclination to view "the sets" as a constituting a given kind rests on a confusion. (2) Granting the first primitive notion, we can take in place of the second primitive notion the notion of an arbitrary subset. For given the notion of an arbitrary subset can arrive at the operation of powerset by applying the set formation operation to the kind of sets consisting of arbitrary subsets of a given set. (3) One might think that the second primitive notion can be founded on the first. But this is a mistake. Given a set X

the first primitive notion allows us to form sets corresponding to any *kind* of elements of X. This will yield the definable subsets of X but it will not yield all arbitrary subsets of X. Of course, the procedure will yield all arbitrary subsets of X if we appeal to the notion of an arbitrary subset in specifying a given kind; for example, if we let A be an arbitrary subset of X and consider the kind of object (or plurality) corresponding to A. But such a justification is circular. The notion of an arbitrary subset is fundamental. The notion of a plurality in a sense richer than that associated with definability is derivative.

1.1.2 The Iterative Hierarchy. Given the above primitives we are now in a position to give an informal characterisation of the iterative hierarchy of sets.

The powerset operation can be iterated. Thus, starting with a given set, say the set of the natural numbers ω, we can take its powerset to obtain $\mathscr{P}(\omega)$ and then we can take the powerset of this in turn to obtain $\mathscr{P}(\mathscr{P}(\omega))$, and so on. In this way we generate a hierarchy ω, $\mathscr{P}(\omega)$, $\mathscr{P}(\mathscr{P}(\omega))$, ... Our starting point in this case was ω. But we could have started with any set. For example, we could have started with the set of giraffes or the set of fundamental particles. Of all possible starting points there is a simplest, namely, the set which contains nothing, that is, the *empty set*, which we shall denote $\varnothing$. Let us start here, at the simplest point, and construct the hierarchy of *pure sets*.

The first level of the hierarchy is $V_0 = \varnothing$. We can take the powerset of this to obtain $V_1 = \mathscr{P}(\varnothing)$. Continuing in this way we obtain $V_2 = \mathscr{P}(\mathscr{P}(\varnothing))$, $V_3 = \mathscr{P}(\mathscr{P}(\mathscr{P}(\varnothing)))$, ... What next? After all of the finite stages there is the first limit stage ω. At this stage we form the set $V_\omega = \bigcup_{n<\omega} V_n$, that is, the set consisting of everything that came before. We are then in a position to start taking powersets again. And so we continue to the next limit stage $\omega + \omega$. What next? At this stage we can form the set $V_{\omega+\omega}$ consisting of everything that came before, that is $V_{\omega+\omega} = \bigcup_{\alpha<\omega+\omega} V_\alpha$. We are then in a position to start taking powersets again ...



Notice that there are three salient features in the above account. First, we applied an operation, the *powerset operation*, to get from one successor stage to the next. Second, we applied an operation, the *summing up operation* (a special case of the set formation operation), to pass through limit stages. Third, we ended our description with *the elusive three dots* ... The first two operations are clear but the three dots are something of a puzzle since they never seem to go away. For example, suppose we have applied the summing up operation ω-many times. Even then we are not at an end. For we can apply the summing up operation to get $V_{\omega\cdot\omega}$. Then we can start taking powersets again ...

1.1.3 Nature. There are two axioms of nature: Extensionality and Foundation.

Extensionality serves to distinguish sets from other objects, such as concepts, which also, in some sense, collect a plurality into a unity. The axiom of Extensionality says that two sets A and B are the same if and only if they have the same members; that is, a set is *determined* by its members. This is in contrast to a concept. For compare the concepts *beam of light* and *array of photons*. These concepts apply to exactly the same things (that is, they have exactly the same things "falling under" them) and yet they are *not* the same; that is, a concept, in contrast to a set, is not determined by the objects that "fall under" it.

Foundation serves to rule out certain pathologies by stipulating that each set b is *grounded* in the sense that if you pick a member of b and then pick a member of *that* and then pick a member of *that* and so on, then you will eventually "ground out" at a set which has no members; that is, Foundation asserts that every non-empty set

has an $\in$-minimal element.

The axioms of Extension and Foundation pertain to the *nature* of sets. Both are implicit in our informal characterization of the concept of set as built up from below by iterating the powerset operation.

1.1.4 Extent. But there is more that is implicit in our informal account than just the axioms of nature and it is here that we encounter the first issue surrounding the elusive three dots, namely, the issue of how far the hierarchy extends. We saw that there is a set $\varnothing$ and when we applied the summing up operation for the first time we obtained an infinite set V_ω. Thus, already at this stage, we have two axioms of extent: the axiom of Emptyset and the axiom of Infinity. Going further we saw that there is a level $V_{\omega\cdot\omega}$. At this point we took the easy way out by writing down three dots.

One would like to say something *general* about the elusive three dots. There are two issues. First, there is the issue of whether the three dots span a *determinate* totality. Second, there is the issue of which levels the three dots span; for example, whether there is a level V_κ that satisfies all of the axioms of ZFC.

1.1.5 Two Conceptions: Actualist and Potentialist. The first issue surrounding the elusive three dots is whether they span a *determinate* plurality. There are two views:

(1) ACTUALIST: The sets form a determinate plurality.

(2) POTENTIALIST: The sets do not form a determinate plurality.

In other words, the actualist maintains that "the sets" form a given kind and the potentialist maintains that they do not.

The actualist owes us an account why the summing up operation does not apply to "the sets". After all, according to the actualist "the sets" are on a par with the finite sets insofar as both are determinate pluralities (or given kinds). What then is it about the former that distinguishes it from the latter in such a way as to render

the set formation operation inapplicable? The actualist can tell us that the former is a 'proper class'. But this doesn't answer our question, it just shifts it. For then we should like an account of the summing up operation does not apply to proper classes. What is it about given kinds which form proper classes that distinguishes them from given kinds that do not in such a way as to render the set formation operation inapplicable?

The potentialist, of course, doesn't face this problem since the potentialist thinks that the summing up operation always applies and that consequently there is no determinate totality V. The actualist finds this absurd. For it would seem that the above account explicates the concept of set. Why then can we not just consider the extension of this concept? Surely this is a determinate totality. The potentialist must maintain that the above account does not explicate *the* concept of set. How? I suggest the following response on behalf of the potentialist: It is a mistake to speak of *the* above account since we concluded it with the elusive three dots. *If* we stop the account at some stage, say when we obtained $V_{\omega \cdot \omega}$, and we treat this as an account of 'the concept of C-set' then we can indeed make sense of the extension of the concept of C-set. But the concept of set goes beyond the concept of C-set; the extension of the concept of C-set *is* a set. What then of *the* concept of set. Well, there is more to being a set than being an extensional well-founded object, just as there is more to being a natural number than having finitely many predecessors. For example, just as it is impossible for there to be a largest natural number n it is impossible for there to be a largest set X. In the first case this is because the concept of natural number involves the closure condition that every number has a successor; in the second case this is because the concept of set involves the closure condition that every set has a powerset. But, in contrast to the case of the concept of natural number, there is no *single* closure condition (such as closure under powerset) which exhausts the concept of set. Every time one tries to give a *definite* account of which closure conditions embody the concept of set one ends up having described a *new* set, namely, the set closed under the proposed closure conditions. One might try to get around this difficulty by characterizing the concept of set in terms of "all" such closure conditions

but again this involves the elusive three dots. Until we have been told what is meant by the elusive three dots we have not fixed the concept of set and so we have not fixed the extension V of the concept of set.

I want to underscore two respects in which the actualist-potentialist distinction in the context of set theory differs from the same distinction in the context of number theory. First, in the context of number theory the actualist could retreat to higher ground, say set theory, to justify the view that the totality of natural numbers is an actual completed totality. But in the context of set theory, the universal nature of the subject robs the actualist of such a retreat. Second, in the context of number theory the potentialist view leads to a system which is quite different from a mathematical point of view than that obtained on the actualist. This is because the central problems of number theory are not about specific numbers, but instead about all of the numbers. Many such problems will be regarded as indeterminate by the potentialist. In set theory the situation is quite different. Almost all of the problems in set theory pertain to an initial segment of the universe (the analogue of a specific number) and on such questions the potentialist and the actualist will agree.

In summary, I have raised two problems, one for the actualist and one for the potentialist. The problem for the actualist is to explain why the operation of set formation does not apply to the determinate totality consisting of "the sets". What is it about the *nature* of this totality (of this given kind) that resists being brought together to form a set? I do not see how to answer this question. The problem for the potentialist is to explain why "the sets" do not form a given kind. I have tried to meet this challenge above by underscoring the fact that "the" concept of set like the concept of natural number involves closure principles, but that in contrast to the concept of natural number there is no fixed set of closure principles that we can point to—each determination of closure principles leads to other richer closure principles, each attempted determination of "the" concept of set falls short.

I do not profess to have settled the issue between the actualist and the potentialist. I don't even have a settled view on the matter. For the moment I would like explore the terrain by investigating the extent to which the actualist and the potentialist can

justify new axioms. This will entail elaborating the potentialist view. One thing that I will show is that a version of the potentialist view provides a justification of axioms which are much stronger than those which can be justified on the actualist view. This will be shown in Chapter 4.

1.1.6 Stage Theory. The second issue surrounding the elusive three dots is which levels they span. We have already seen in §1.1.4 that the levels V_ω and $V_{\omega\cdot\omega}$ exist. Is there a level V_κ which is a model of ZF?

One of the traditional ways of attempting to turn our informal account into a more precise account—one which will enable us to see more clearly which levels exists—is due to Shoenfield (1967). Here is a description of the account:

> When can a collection of sets be formed into a set? For each set x in the collection , let S_x be the stage at which x is formed. Then we can form a set of this collection iff there is a stage S which follows all the S_x. ... Since we wish to allow a set to be as arbitrary a collection as possible, we agree that there shall be such a stage whenever possible, i.e. whenever we can visualize a situation in which all of the stages in the collection are completed. (Shoenfield (1967, p. 238–9))

In order for this account to be precise one should have to explain the notion of possibility. Shoenfield does this in terms of what is visualisable. But this will not do since most of the stages that occur in Shoenfield's discussion (for example, $V_{\aleph_1}$) are not visualisable; furthermore, such stages (as Parsons (1977) argues) are not intuitable in any reasonable sense of the notion.

Shoenfield does give an example of what is possible or visualisable in the sense he intends—the *principle of cofinality*:

> Suppose that we have a set A, and that we have assigned a stage S_a to each element a of A. Since we can visualize the collection A as a single object (viz. the set A), we can also visualize the collection of stages S_a as

a single object; so we can visualize a situation in which all of these stages are complete. (Shoenfield (1967, p. 239))

Using the principle of cofinality Shoenfield gives a justification of Replacement. It is hardly surprising that one can do this since the principle of cofinality just is the axiom of Replacement.

What we seek is a general principle which gives a uniform account of a vast array of the axioms of set theory, one which in justifying a given axiom does not presuppose that very same axiom. A good example of such an account is due to Gödel.

1.2 Gödel's Approach

1.2.1 The Account of 1933. Gödel arrives at the iterative conception by liberating the simple theory of types from three restrictions; first, he permits mixed types, second, he treats '$a \in b$' as false (as opposed to meaningless) when the type of b is less than or equal to the type of a, and third, he admits types of transfinite order.[1] Thus, starting with the empty set of individuals the sets of finite type are: $V_0 = \varnothing$, $V_1 = \mathscr{P}(V_0)$, $V_2 = \mathscr{P}(V_1) \cdots V_{n+1} = \mathscr{P}(V_n) \cdots$ The hierarchy then continues:

> You can make the class of all classes of finite type play the role of the class of individuals, i.e., take is as a basis for a new hierarchy of types and thus form classes of type $\omega + 1$, $\omega + 2$, and so on for each transfinite ordinal. (Gödel (*1933o, p. 7))

That is, we have $V_{\alpha+1} = \mathscr{P}(V_\alpha)$, $V_\lambda = \bigcup_{\alpha<\lambda} V_\alpha$ for λ a limit ordinal, and $V = \bigcup_{\alpha\in\Omega} V_\alpha$.[2]

This approach resembles the standard approach in that the ordinals are assumed to be given antecedent to set theory. Gödel makes precisely this point before turning to a refinement of the approach:

> [I]n order to state the axioms for a formal system, including all the types up to a given ordinal α, the notion of this ordinal α has to be presupposed

> as known, because it will appear explicitly in the axioms. On the other hand, a satisfactory definition of the transfinite ordinals can be obtained only in terms of the very system whose axioms are to be set up. (Gödel (*1933o, p. 8))

But we can bootstrap.

> The first two or three types already suffice to define very large ordinals. So you can begin by setting up axioms for these first types, for which purpose no ordinal whatsoever is needed, then define a transfinite ordinal α in terms of these first few types and by means of it state the axioms for the system, including all classes of type less than α. (Call it S_α.) To the system S_α you can apply the same process again, i.e., take an ordinal β greater than α which can be defined in terms of the system S_α and by means of it state the axioms for the system S_β including all types less than β, and so on. (Gödel (*1933o, pp. 8–9))

In this manner we can bootstrap our way up the cumulative hierarchy: At stage ω we sum up to get V_ω. Then we apply the powerset operation to get $V_{\omega+1}$, $V_{\omega+2}$ and so on. Given $V_{\omega+2}$ we can set up a system $S_{\omega+2}$ and define a large ordinal α_1 along which we can iterate the powerset operation to obtain V_{α_1}. We can then set up a system S_{α_1} and define a larger ordinal α_2 which enables us to to bootstrap up to V_{α_2} and so on until we reach a limit stage, at which point we sum up to get V_{α_ω} (where $\alpha_\omega = \sup_{n<\omega} \alpha_n$) as we did at stage $\omega \ldots$

One of the things that needs to be spelled out in this account is the resources at our disposal when defining large ordinals. We are told to set up, in the first step, the system $S_{\omega+2}$ and then define a transfinite ordinal α_1 in terms of $V_{\omega+2}$. Our question is: What role does $S_{\omega+2}$ play? There are four issues: (1) Is Gödel demanding that α_1 be given by a well-ordering which is *provably* a well-ordering in $S_{\omega+2}$? (2) Or can we consider well-orderings which are definable over $V_{\omega+2}$ in the language $\mathscr{L}_{\omega+2}$ of $S_{\omega+2}$? (3) Or, even better, can we consider well-orderings which are definable over $V_{\omega+2}$ *with*

parameters in $V_{\omega+2}$ in the language $\mathscr{L}_{\omega+2}$? (4) Finally, can we appeal to the Axiom of Choice?

It does seem that Gödel is referring to definability with parameters, otherwise one couldn't get very far. And it is clear that he is assuming AC since later he notes that one of the weak points of his account "is connected with the axiom of choice". (Gödel *1933, 19) Granting this, one can obtain pretty long well-orderings. For example, starting with the parameter $V_{\omega+1}$ we can get a well-ordering of length $\beth_1 = 2^{\aleph_1}$, where $\beth_\alpha$ is defined to be $\mathrm{Card}(V_{\omega+\alpha})$, the cardinality of $V_{\omega+\alpha}$. It is more convenient, however, to take a well-ordering of length $\beth_1 + 1$ since this will take us up to the successor stage $V_{\beth_1+1}$ and we can use the parameter $V_{\beth_1}$ to get a well-ordering of length $\beth_{\beth_1} + 1$. After ω-many iterations we reach the first $\beth$-fixed point, that is, the least κ such that $\beth_\kappa = \kappa$. We can then apply the summing up operation to get V_κ. And applying the powerset operation a couple of more times we can get to $V_{\kappa+2}$ and, taking $V_{\kappa+1}$ as a parameter, get an even larger ordinal. So the process doesn't bog down at a fixed point. We can keep going. Notice that:

> There are two different ways of generating types, the first consisting of going over from a given type to the next one and the second in summing up a transfinite sequence of given types, which we did, e.g., in forming the type ω. (Gödel (*1933o, p. 9a))

The second operation (when conjoined with operation of taking powersets) is what ensures that we can keep going. We applied it above to get V_ω, V_{α_ω}, and V_κ. The only other operation we appealed to was the trivial process of taking power sets and the significant operation of bootstrapping. Appearances to the contrary, by the first operation Gödel must be referring to what we have called the 'bootstrapping' operation since this is the centre piece of his discussion. This operation probably also subsumes the operation of 'jumping' from one level to the next by the application of the powerset operation. Nevertheless it is useful to separate three ways of generating types: Powerset, Summing Up, and Bootstrapping. Given that we can "always" apply the powerset operation and the summing up operation there is *no end to the*

hierarchy. But this doesn't tell us very much about *how far* the hierarchy extends. The first instance of bootstrapping, namely, going from $\beth_\alpha + 1$ to $\beth_{\beth_\alpha} + 1$ *does* tell us something about the extent of the hierarchy. Of course, as before there is *no end to the applications of this operation.* But this doesn't add much to the first few applications of the operation: It just gives "more of the same". What we would like is a *new* operation, one which *transcends* the previous operations.

In *1933o Gödel does not elaborate on how one might generate new operations. In fact, he follows the above quoted statement with the following:

> Now the statement made by the axioms of the theory of aggregates is essentially this, that these two processes do not lead you out of the system if the second process is applied only to such sequences of types as can be defined within the system itself. [That is to say: If M is a set of ordinals definable in the system and if to each ordinal of M you assign a type contained in the system, then the type obtained by summing up those types is also in the system]. (Gödel (*1933o, p. 9a))

By 'axioms of the theory of aggregates' Gödel means the axioms of set theory "as presented by Zermelo, Frankel, and von Neumann" (Gödel *1933o, 4), that is, ZF. Thus Gödel is recasting the axioms of ZF as saying that the universe of sets is closed under the two processes that he has introduced.

1.2.2 The Account of 1951. There are two questions that arise: (1) What exactly is the 'second process'? Gödel has given us a hint but he has neither fully specified it nor even specified it to the extent that it clearly yields the axiom of Replacement. (2) What is meant by 'defined within the system'? The answers to both of these questions is implicit in Gödel (*1951), which I will quote at length.

> If, for example, we begin with the integers, that is, the finite sets of a special kind, we have at first the sets of integers and the axioms referring to them (axioms of the first level), then the sets of sets of integers with their axioms (axioms of the second level), and so on for any finite iteration

of the operation 'set of'. Next we have the set of all these sets of finite order. But now we can deal with this set in exactly the same manner as we dealt with the set of integers before, that is, consider the subsets of it (that is, the sets of order ω) and formulate axioms about their existence. Evidently this procedure can be iterated beyond ω, in fact up to any transfinite ordinal number. So it may be required as the next axiom that the iteration is possible for *any* ordinal, that is, for any order type belonging to some well-ordered set. But are we at an end now? By no means. For we have now a new operation of forming sets, namely, forming a set out of some initial set A and some well-ordering B by applying the operation 'set of' to A as many times as the well-ordered set B indicates. And, setting B equal to some well-ordering of A, now we can iterate this new operation, and again iterate it into the transfinite. This will give rise to a new operation again, which we can treat in the same way, and so on. So the next step will be to require that *any* operation producing sets out of sets can be iterated up to any ordinal number (that is, order type of a well-ordered set.) But are we at an end now? No, because we can require not only that the procedure just described can be carried out with any operation, but that moreover there should exist a set closed with respect to it, that is, one which has the property that, if this procedure (with any operation) is applied to the elements of this set, it again yields elements of this set. You will realize, I think, that we are still not at an end, nor can there ever be an end to *this* procedure of forming the axioms, because the very formulation of the axioms up to a certain stage gives rise to the next axiom. (Gödel (*1951, pp. 3–5))

This account coincides with and extends that of *1933o with the slight difference that Gödel starts with $V_0 = \mathbb{Z}$. For comparison let us start with $V_0 = \varnothing$. Thus, as before Gödel starts with successive applications of the power set and summing up operation. This gives V_0, $V_1 \cdots V_n \cdots V_\omega$, $V_{\omega+1}$, $V_{\omega+2} \cdots$ The final three dots here indicates "more of the same". Once we are at some fixed level, say $V_{\omega+2}$ we can define various

ordinals α_1+1 and iterate the powerset operation α_1+1 many times to get V_{α_1+1}. Call this an instance of the first jump operator. Next we consider the first jump operator in the general context. This is the operator which given a well-ordering of a set A, say $|A|$, allows us to iterate the powerset operation $|A|$-many times over A, that is, the operation sending A to $\mathscr{P}^{|A|}(A)$. As before, starting with the parameter $V_{\omega+1}$ we can get a well-ordering of length $A = \beth_1 + 1$ which will bootstrap us up to $V_{\beth_1+1}$ and then we can use the parameter $V_{\beth_1}$ to get a well-ordering of length $\beth_{\beth_1} + 1$ and so on. After ω-many iterations we reach the first $\beth$-fixed point, that is, the least κ such that $\beth_\kappa = \kappa$. This is where the discussion of *1933o ended. But in *1951 Gödel continues. We can *reflect* on the previous procedure to consider the operation that sends α to the α^{th} $\beth$-fixed point. Call this an instance of the second jump operator. The second jump operator transcends the first. And we can continue to apply it to yield "more of the same". But as before it can be transcended to yield a third jump operator and so on. In this way we yield richer and richer jump operators all of which map sets to sets. The next step is to reflect on all of this.

> But are we at an end now? No, because we can require not only that the procedure just described can be carried out with any operation, but that moreover there should exist a set closed with respect to it, that is, one which has the property that, if this procedure (with any operation) is applied to the elements of this set, it again yields elements of this set. (Gödel (*1933o, p. 4))

That is, we can consider the operator which maps and ordinal α to the least V_κ closed under all of the set functions. This answers both of our above questions. First, Gödel has unfolded more of what he means by the second process. Second, we see that by 'defined within the system' Gödel did not mean definable over the background universe V; he must have meant that (as he says in *1951) that we are to apply the operations in V to sets in V. Further evidence of this second point can be found in the following statement that Gödel certified:

> From the very idea of the iterative concept of set it follows that if an

> ordinal number α has been obtained, the operation of power set ($\mathscr{P}$) iterated α times leads to a set $\mathscr{P}^{\alpha}(0)$. But, for the same reasons, it would seem to follow that if, instead of $\mathscr{P}$, one takes some larger jump in the hierarchy of types, e.g. the transition $\mathscr{Q}$ from x to $\mathscr{P}^{|x|}(x)$ (where $|x|$ is the smallest ordinal of the well-orderings of x), $\mathscr{Q}^{\alpha}(0)$ likewise is a set. Now, to assume this for any conceivable jump operation (even for those that are defined by reference to the universe of all sets or by the use of the choice operation) is equivalent to the axiom of replacement. (Wang (1974, p. 186))

It is not too hard to see that a universe which is closed under *all* such operations from sets to sets is a model of ZF. Thus we have a justification of the axiom of Replacement. Even more, one can continue, as Gödel notes in the above quoted passage from *1951 and the following continuation of *1933o:

> But it would be a mistake to suppose that with this system of axioms [namely, ZF] for the theory of sets we should have reached an end to the hierarchy of types. For all the classes occurring in the system can be considered as a new domain of individuals and used as a starting point for creating still higher types. There is no end to this process [and the totality of all systems thus obtained seems to form a totality of similar character to the set of ordinals of the second ordinal class]. (Gödel (*1933o, pp. 9–10))[3]

In this way one obtains a proper class of inaccessible cardinals.

Note that Gödel's approach proceeds from the bottom up. This is an advantage. But there are two weaknesses in the account. First, it presupposes the Axiom of Choice. Second, it cannot yield weakly compact cardinals. One way to overcome both of these weaknesses is to resort to a top down approach. Gödel actually thought that such an approach would prove fruitful. In fact he was rather optimistic. In 1977 he wrote:

> Generally I believe that, in the last analysis, every axiom of infinity should be derivable from the (extremely plausible) principle that V is indefinable, where definability is to be take in [a] more and more generalized and idealized sense. (Wang (1977, p. 325); Wang (1996, p. 285))

One way of making the statement 'V is undefinable' precise is in terms of reflection principles, that is, principles which assert that if a formula with parameters holds in V then it holds (with the parameters relativised) in some rank initial segment V_α. These principles get around both of the above difficulties, as will be shown in §1.3. The following (somewhat technical) section on AC can be skipped without loss of continuity.

1.2.3 Choice. The axiom of choice does not neatly fit into our division of the axioms into those of nature and those of extent. There has been a great deal of controversy both over the truth and the nature of the axiom. In this section I will restrict myself to a couple of brief remarks.

It has been argued by some (for example, Tait) that AC is a truth of second-order logic which follows from the meaning of the quantifiers. This strikes me as doubtful for the following reason: A cardinal κ is *Reinhardt* iff there is a (non-trivial elementary) embedding $j : V \to V$ such that $\mathrm{cp}(j) = \kappa$. Kunen showed assuming ZFC that there are no Reinhardt cardinals. But it is open whether assuming ZF there can be Reinhardt cardinals. This is a serious question. It amounts to the question of whether the canonical hierarchy of consistency strength outstrips ZFC. It is not the sort of question which can be ruled out on the basis of logic alone. But according to the view that AC is a principle of logic it *can* be ruled out by logic alone.

It has also been argued (for example, by Parsons) that our reasons for believing AC are a posteriori. This seems right but it has a surprising consequence. For it may be the case that the hierarchy of consistency strength outstrips that of the large cardinals consistent with Choice. Furthermore, these axioms might be "so abundant in their verifiable consequences, [shed] so much light upon a whole discipline, and [furnish] such powerful methods for solving given problems (and even solving them, as far

as that is possible, in a constructive way) that quite irrespective of their intrinsic necessity they would have to be assumed at least in the same sense as any well-established physical theory" (Gödel (1964, p. 265)). In such a situation we would have to abandon AC. One could view AC as a limitative principle on a par with $V = L$. Here $V^{\#}$ is to AC what $0^{\#}$ is to $V = L$.

1.3 Reflection Principles

1.3.1 Motivation. Reflection principles aim to make precise the idea that V is undefinable, an idea that played a central role in Cantor's original introduction of set theory and in the subsequent accounts of Zermelo, Ackermann, Bernays, Gödel, and Reinhardt. On this view V is taken as an additional primitive and the axioms pertaining to it are axioms of nature.

Before giving a precise formulation of reflection principles let me motivate the idea that V is undefinable. Recall that in §1.1.5 I argued that the concept of set involves closure principles. Well, which closure principles? It is useful to extract approximations to the full concept of set by selecting a definite collection of closure principles C. For example, in a very simple case, C might be *closure under powerset.* Let us refer to this approximation of the concept of set as the *concept of C-set.* A useful notion is the *spectrum of* C, defined to be $\Lambda(C) = \{\alpha \mid V_\alpha \text{ is closed under } C\}$. For each V_α, where $\alpha \in \Lambda(C)$, we shall say that V_α *meets* the concept of C-set. We shall also say that V_α is a *permissible interpretation* of the concept of C-set. Thus, in our present example the first level which meets the concept of C-set is V_ω, the second is $V_{\omega+\omega}$, and in general $\Lambda(C)$= Limit Ordinals. By laying on the closure conditions, one thins out the spectrum and approaches the full concept of set. Notice that as soon as we fix some *definite* set of closure conditions C and we consider the precise concept of C-set we immediately see that it falls short of the concept of set. This is a key feature of *the* concept of set: it resists characterisation in terms of some definite set of closure principles. Thus V is "uncharacterisable" and "undefinable".

1.3.2 Formulation. The standard way of making the idea that V is undefinable more precise is in terms of *reflection principles.* These principles assert that any statement true in V is true in some smaller V_α. Thus, for any φ one cannot define V as the collection which satisfies φ since for any such φ there will be a proper initial segment V_α of V that satisfies φ. More formally, we shall write this as

$$V \models \varphi(A) \;\rightarrow\; \exists\alpha\; V_\alpha \models \varphi^\alpha(A^\alpha)$$

where $\varphi^\alpha(\,\cdot\,)$ is the result of relativising the quantifiers of $\varphi(\,\cdot\,)$ to V_α and A^α is the result of relativising an arbitrary parameter A to V_α. This is merely a preliminary characterization of a reflection principle since we have left open (i) the specification of the language, (ii) the order of parameters, and (iii) the nature of relativisation. We will fill in these three parameters as we proceed and thereby arrive at the general characterization of a reflection principle.

Language: At the most general level our language will be $\mathscr{L}_{\beta,\gamma}$, the language of set theory with quantifiers of order $\leq \beta$ and parameters of order $\leq \gamma$. I will abbreviate $\mathscr{L}_{\beta,\beta}$ as $\mathscr{L}_\beta$. When the quantifiers are to be strictly less than β I will write $\mathscr{L}_{<\beta,\gamma}$ and likewise for parameters. I will use lower case letters $x, y, z, \ldots$ to range over objects of first order and upper case letters $X^{(\beta)}, Y^{(\beta)}, Z^{(\beta)}, \ldots$ to range over objects of β^{th}-order and I will drop the superscripts the order is clear from context. The distinction between first- and higher-order objects and formulas will always be relative to a fixed universe of sets; thus if our universe of first-order objects is V_κ we regard $X^{(\beta)}$ as ranging over $V_{\kappa+\beta-1}$ if β is a successor ordinal, and over $V_{\kappa+\beta}$ if β is a limit ordinal. The order of a formula is taken to be the maximum of the order of its bound variables.

The *Levy hierarchy* of formulas is defined as follows: A sentence is $\sum_n^\beta$ if it is of the form $\exists x_1 \forall x_2 \cdots Q x_n\, \varphi(x_1 \cdots x_n, a_1 \cdots a_m)$ where Q is $\forall$ if n is even and $\exists$ if n is odd and, as indicated by the superscript in $\sum_n^\beta$, the quantifiers and parameters are of order $\leq \beta$. When we want to restrict the parameters to those of a given order, say those of order $\leq \gamma$ ($< \gamma$), we will write 'γ' ('$< \gamma$') after 'β'. For example, $\Pi_n^{\beta,\gamma}$ consists of formulas in Π_n^β with parameters of order $\leq \gamma$. The fact that we are

allowing parameters is notationally indicated by the occurrence of '$\sim$' underneath 'Σ'. When the '$\sim$' under 'Σ' is absent this indicates that parameters are not allowed. $\underset{\sim}{\Pi}^{\beta}_{n}$ sentences are negations of $\underset{\sim}{\Sigma}^{\beta}_{n}$ sentences.

Relativisation: In general, we will deal with parameters of *arbitrary* order but in the present chapter we will restrict ourselves to parameters of *second* order, that is, classes in the standard sense. We relativise second-order classes to a level V_α by omitting elements in the class which are not in V_α; more precisely, $A^{(2)\alpha} = A^{(2)} \cap V_\alpha$ is the relativisation of the second-order class $A^{(2)}$ to V_α. This is how someone living in V_α views the class. We relativise formulas to a level V_α by restricting the bound variables to the appropriate level; more precisely, the relativisation of $\varphi(A_1^{(\beta_1)} \dots A_n^{(\beta_n)})$ to V_α is the statement $\varphi^\alpha(A_1^{(\beta_1)\alpha} \dots A_n^{(\beta_n)\alpha})$ obtained by bounding the β^{th}-order quantifiers with $V_{\alpha+\beta-1}$ if β is a successor, and with $V_{\alpha+\beta}$ if β is a limit, relativising the class parameters to sets as indicated. This is how someone living in V_α interprets the formula.[5] To distinguish notationally between classes and sets we will use $[\cdots]$, $[x \mid \varphi(x)]$ and the like in place of $\{\cdots\}$, $\{x \mid \varphi(x)\} \dots$ and if we wish to be explicit about the order of a class we will append a superscript such as '(β)' as in the case of variables and parameters.

1.3.3 First Order. Given the above notation we are now in a position to give a more general formulation of reflection principles. Let Γ be a collection of formulas in $\mathscr{L}_{\beta,2}$. V is Γ-*reflective* iff

$$\forall X \left(\varphi(X) \to \exists \alpha \, \varphi^\alpha(X^\alpha)\right)$$

for all $\varphi \in \Gamma$ where X is of second order. With the exception of the restriction to second-order parameters with scheme is fully general. The discussion of third and higher order parameters, which is delicate, will be taken up in §3.2 after we have exhausted the resources of reflection with second-order parameters.

We are now in a position to begin mapping out the extent of reflection. I will first state the results in the first order context and then restate and elaborate on these results in the second-order context.

Let T be ZF−Replacement−Infinity. This is the theory that we have justified above and our aim is to justify ZF.

FACT 1. *Assume* T + $\mathscr{L}_{1,1}$-reflection. *Then* ZF *holds.*

We would like to reflect on the above fact to obtain a level V_α which is a model of ZF. The obstacle to doing this in the context of $\mathscr{L}_{1,1}$ is that the theory ZF is not finitely axiomatisable and so there is no single sentence that we can reflect. In fact, the theory T + $\mathscr{L}_{1,1}$-reflection is equivalent to ZF. But there is a schematic version of ZF and a schematic version of $\mathscr{L}_{1,1}$-reflection (which I will call ZF′ and $\mathscr{L}'_{1,1}$-reflection) such that we have the following:

FACT 2. *Assume* ZF′ + $\mathscr{L}'_{1,1}$-reflection. *Then there is a proper class of inaccessible cardinals.*

(See Feferman (1996) for details.)

1.3.4 Second Order. In this subsection I want to give the reader some insight into the manner in which reflection principles justify the axioms of extent. It will be convenient to work in $\mathscr{L}_{2,2}$.

Minimal set theory (MST) consists of the following axioms of nature: Extensionality, Second-order Separation, Foundation and the obvious axioms governing 'rank'. MST was introduced in Tait (1998b). Note that MST says *nothing* about the extent of the set theoretic hierarchy; its models are exactly the rank initial segments V_0, V_1 ... V_α ... Given a theory T, let Spec(T) = $\{\alpha \mid V_\alpha \models T\}$ be the *spectrum of* T. For example, Spec(MST) = Ω (where 'Ω' is the name for the totality of ordinals) is another way of expressing the fact that MST makes no demand on the height of the universe. Our aim is to make such demands by supplementing MST with reflection principles. This will have the effect of "thinning out" the spectrum.

The first step is to derive standard set theory with reflection principles. (The method of doing this is by now part of the folklore of the subject but I believe it can be traced back to early work of Ackermann. I will follow the exposition of Tait (1998b).) Work in MST.

(1). Consider the formula $\varphi_1(X)$ asserting that the second-order class X is coextensive with a set. Since V is undefinable it is not defined by φ_1. In terms of reflection this says

$$\forall X \left(\varphi_1(X) \rightarrow \exists \alpha \, {\varphi_1}^{\alpha}(X^{\alpha})\right),$$

that is, if X is coextensive with a set from the point of view of V, then it is coextensive with a set from the point of view of V_α. This can happen if and only if V is a limit level. Thus, Spec(MST + φ_1-reflection) = Limit Ordinals. Notice also that in MST + φ_1-reflection we can derive the standard axioms of Union, Pairing and Powerset and we can *formulate* AC.

(2). Consider the sentence φ_2 asserting that there are sets of arbitrarily high rank. Since V is undefinable it is not defined by φ_2. In terms of reflection this says

$$\forall X \left(\varphi_2(X) \rightarrow \exists \alpha \, {\varphi_2}^{\alpha}(X^{\alpha})\right),$$

that is, if V thinks that there are sets of arbitrarily high rank the some smaller level V_α must think that there are sets of arbitrarily high rank. But V *does* think this, something we established by φ_1-reflection. So there is a level V_α which thinks that there are sets of arbitrarily high rank. Thus, Spec(MST + φ_1, φ_2-reflection) = Limit Ordinals$-\{\omega\}$ and this theory is equivalent to ZF$-$Replacement.

(3). Consider the sentence $\varphi_3(X)$ asserting that X codes a function whose domain is a set. (Note that we are saying nothing about whether its range is a set.) Since V is undefinable it is not defined by φ_3. In terms of reflection this says

$$\forall X \left(\varphi_3(X) \rightarrow \exists \alpha \, {\varphi_3}^{\alpha}(X^{\alpha})\right),$$

that is, if V thinks that X is a code for a function whose domain is a set then so does some smaller V_α. But the domain, being a set, gets reflected to itself in V_α and so the range, which hasn't changed either, is a class of V_α, that is, a *set* of V. Thus, reflecting on φ_3 has the effect of ensuring that every function applied to a set yields a set. This is just the Replacement axiom. Thus MST + $\varphi_1, \varphi_2, \varphi_3$-reflection

is just ZF and Spec(MST + $\varphi_1, \varphi_2, \varphi_3$-reflection) = Inaccessibles. Notice also that ZF = MST + $\mathscr{L}_{1,2}$-reflection. Thus, by supplementing MST with reflection principles we have derived ZF.

(4). Consider the sentence φ_4 asserting that height of the universe is inaccessible. By (3) this is true of V. And so by reflection there must be an inaccessible cardinal. Let φ_5 be the sentence which asserts that the universe has inaccessible height *and* that it contains an inaccessible cardinal. So this is true of V. So by reflection there are *two* inaccessible cardinals. Continuing in this way we obtain a proper class of inaccessible cardinals.

1.3.5 Conclusion. In this chapter I considered three approaches to justifying the standard axioms of set theory: the traditional approach, Gödel's approach, and the approach based on reflection principles. The traditional approach is unsatisfactory since it is circular. Gödel's approach is unsatisfactory since it assumes AC. The approach in term of reflection principles is free from both of these problems and has three additional virtues: First, it provides a uniform account of all of the standard axioms of extent. Second, it has shown promise in being able to justify axioms beyond ZF. Finally, it partially fleshes out one of the central features of the universe of sets, namely, that it is uncharacterisable.

Our goal now is to determine the extent of reflection principles and connect this to our ultimate aim of reducing the limitations of ZF. We should like to know whether reflection principles can effect a "significant" reduction in incompleteness. But for this we must make the notion of a "significant reduction" mathematically precise. This will be the goal of the next chapter.

Chapter 2

Independence

Our goal is to find and justify an extension of ZF which effects a "significant" reduction in incompleteness. The purpose of this chapter is to analyse the limitations of ZF and give a mathematically precise formulation of the notion of a "significant" reduction in incompleteness. This will put us in a position to prove theorems of the form: Axioms of type A cannot yield a significant reduction in incompleteness.

2.1 Limitations

2.1.1 The Incompleteness Theorems. Let us say that a theory T is *sufficiently strong* if it is strong enough to encode syntax. For example, PA and the much weaker theory Q are sufficiently strong. According to Gödel's first incompleteness theorem any sufficiently strong recursively enumerable theory T is incomplete. According to Gödel's second incompleteness theorem no sufficiently strong recursively enumerable theory can prove its own consistency. (Henceforth all of the theories we will consider will be sufficiently strong and recursively enumerable.)

Suppose that T is consistent. Then there is a *true* statement that T cannot capture, namely, Con(T). Thus the theories Q, PA, and ZF fall short of capturing truth, as witnessed by the true statements Con(Q), Con(PA) and Con(ZF). Let us concentrate on ZF.

There is an obvious way to overcome this particular limitation of ZF, namely,

add the true sentence Con(ZF). The trouble is that the new system ZF + Con(ZF)—although able to capture the truth Con(ZF)—is unable to capture the truth Con(ZF + Con(ZF)). But we can fix this by extending our system to ZF + Con(ZF) + Con(ZF + Con(ZF)). In this way we obtain a series of systems each one of which captures a truth missed by its predecessor.

2.1.2 Benign Limitations. There is something benign about the above brand of incompleteness. One reason is that if we have good reason for believing a theory T then we have good reason for believing the statement Con(T) witnessing the incompleteness of T.[4] More importantly, if we know that Con(ZF) is a limitation then we know whether or not Con(ZF) is true. I am not saying that all consistency statements are benign; for example, there is nothing benign about Con(ZF+'There is a supercompact cardinal'). Rather, I am saying that consistency statements that *we know to be instances of incompleteness* are benign. The reason such statements are benign is that if you know that a consistency statement φ (or indeed any Π^0_1-statement) is independent of a correct theory T (which includes a minimal amount of arithmetic, such as Q) then you know that φ is true. (Proof: Let $\varphi = \forall n\, \psi(n)$ where ψ is a bounded formula. Suppose, for contradiction, that φ is not true. Then $\exists n \neg\psi(n)$. Then $\neg\psi(k)$ for some k. But then $\mathrm{Q} \vdash \neg\psi(k)$ since Q is correct for bounded statements. So $\mathrm{Q} \vdash \exists n \neg\psi(n)$, which contradicts the assumption that φ is independent of T). The point is that we do not know that Con(ZF+'There is a supercompact cardinal') is an instance of incompleteness, since we don't know that it is true.

To summarise: We are discussing the *limitations* of ZF, that is, we are considering statements φ which are independent of ZF. So far we have encountered one kind of limitation, namely, the consistency statements. These limitations are *benign* in that if we know that a consistency statement φ is independent of ZF then we know that φ is true.

Are all limitations benign?

2.1.3 Serious Limitations. No. Combined results of Gödel and Cohen provide a vast array of serious statements which we know to be independent of ZF without knowing whether or not they are true. Given what we have said above these statements are necessarily not Π^0_1. The classic example of such a statement is CH, a Σ^2_1-statement.

Let me summarise the contrast between benign and serious limitations. The statement Con(ZF) is independent of ZF. It exemplifies a limitation of ZF. But it is benign because the fact that it is a limitation implies that it is true. This is the case with all Π^0_1-statement. The statement CH is independent of ZF. But it is serious because the fact that it is a limitation does not imply that it is true; indeed the fact that it is a limitation does not provide one shred of evidence that it is true.

The serious limitations pose serious problems. For they put us in a situation where we know that there are statements that our system is not capturing and yet do not know whether or not these statements are true. The challenge for new axioms is to overcome the serious limitations.

2.1.4 Example. CH is an example of a serious limitation but as we will see in Chapter 5 it is a *serious* serious limitation. I want to give a more modest example of a serious limitation.

SPHERE. *One cannot take a sphere, split it into finitely many $\underset{\sim}{\Sigma}^1_2$ pieces, rearrange those pieces via rigid motions and put them back together to form a sphere of twice the size.*

This is a $\underset{\sim}{\Sigma}^1_3$-statement which is independent of ZF. It is a serious limitation—we know that either it or its negation exemplifies a limitation of ZF and yet this knowledge does not provide us with the answer to whether or not it is true.

A couple of remarks are in order. First, if one drops the restriction that the pieces be $\underset{\sim}{\Sigma}^1_2$ then, assuming AC, there are very bizarre pieces which render the sentence is false, as Tarski and Banach demonstrated. Second, if one replaces the restriction that the pieces be $\underset{\sim}{\Sigma}^1_2$ by the restriction that they be $\underset{\sim}{\Sigma}^1_0$ or $\underset{\sim}{\Sigma}^1_1$, then the statement is actually

provable in ZFC. In short, there is no paradoxical decomposition involving simply definable (meaning $\underset{\sim}{\Sigma}^1_1$) pieces but there is one involving complicated undefinable pieces. And it is independent of ZF whether there is a paradoxical decomposition involving $\underset{\sim}{\Sigma}^1_2$ pieces.

So a particular challenge for new axioms is to resolve SPHERE.

2.2 The Minimal Hurdle

Our goal is to overcome the serious limitations of ZF by searching for new axioms. But we have to be careful. On the one hand, we don't want to set the bar too high and try to overcome too many of the serious limitations in one shot. On the other hand, we don't want to set the bar too low and overcome a mere isolated instance of a serious limitation, such as the limitation involving SPHERE. The aim of the present section is to analyze the serious limitations of ZF and locate a reasonable place to set the bar. I will start by introducing some technical machinery. Using this machinery I will describe the epistemic situation regarding the serious limitations of ZF. I will then isolate the mathematical source underlying the epistemic situation and use it to set up a mathematically precise hurdle—the Minimal Hurdle—that any extension of ZF must clear if it is to effect a significant reduction in incompleteness.

2.2.1 Degrees of Interpretability. Given a recursively enumerable theory T we can find in an effective manner a primitively recursive theory T′ that has the same provable consequences. This result—due to Craig—allows us to focus, without loss of generality, on primitive recursive theories. Suppose that S and T are primitive recursive theories. We write $S \leq T$ iff there is a primitive recursive translation function τ from from the language of S into the language of T which is such that if $S \vdash \varphi$ then $T \vdash \tau(\varphi)$. In such a situation we say that S is *interpretable* in T. (For a definition of 'translation function' and further details regarding the material in this subsection see Lindström (1997, Ch. 6–8)).

Let D be the set of equivalence classes induced by $\leq$ and let $\mathscr{D} = (D, \leq)$ where

$\leq$ is treated as the induced relation on degrees and all theories extend ZF. $\mathscr{D}$ is the set of *degrees of interpretability* with respect to ZF.

There are three important features of $\mathscr{D}$. First, it is a distributive lattice; in particular, it contains incompatible elements. It is a striking fact that all natural theories appear to be *linearly* ordered in $\mathscr{D}$. Second, $S \leq T$ iff $S \vdash_{\Pi^0_1} \subseteq T \vdash_{\Pi^0_1}$, where $T \vdash_{\Pi^0_1}$ is the set of Π^0_1 consequences of T. Third, $S \leq T$ iff $T \vdash \mathrm{Con}(S|n)$ for all n.

The second of these features implies that $\mathrm{ZF} + \mathrm{Con}(\mathrm{ZF}) > \mathrm{ZF}$. The third implies that $\mathrm{ZF} + \mathrm{CH} \leq \mathrm{ZF}$. This is the crucial difference between the independence technique—*the method of Gödel*—which demonstrates that Con(ZF) is independent of ZF and the independence technique—*the method of forcing*—which demonstrates that CH is independent of ZF. In the former case one moves upward in the degrees of interpretability; in the latter case one does not move at all.

2.2.2 The Scope of ZF. In order to motivate the Minimal Hurdle let us examine in an informal light the successes and failures of ZF. Let's work our way up V and locate the point at which the serious limitations of ZF first arise. The finite levels $V_0, V_1, V_2, \ldots$ are no problem: indeed there are *no* limitations of ZF with regard to these levels. The first infinite level, V_ω, is essentially the natural numbers $\mathbb{N}$, and the set of sentences true in this structure is known as *first order arithmetic.* There are sentences φ of first order arithmetic known to be independent of ZF, for example, Con(ZF). But this is a benign example. What about serious limitations? Surprisingly, there are no known natural serious limitations.

FACT 3. *There is no known example of a natural sentence φ of first order arithmetic such that (i) φ is known to be independent of* ZF *and (ii) it is not known whether or not φ is true.*

I should stress that this is an empirical fact, not a theorem.

The next infinite level, $V_{\omega+1}$, is essentially the real numbers $\mathbb{R}$ and the collection of sentences true in this structure are known as *second-order arithmetic.* It is already here, at the second infinite level—which is just the second infinite rung on the

transfinite ladder—that ZF falters. In fact, it falters before even getting close to this rung. To explain this recall that we split the language of second order arithmetic into increasing levels which we denoted as: Σ^1_1, Σ^1_2, Σ^1_3 ... Corresponding to these language fragments we have fragments of second-order arithmetic. Surprisingly, there are no known natural serious limitations of ZF with respect to Σ^1_2.

FACT 4. *There is no known example of a natural Σ^1_2 sentence φ such that (i) φ is known to be independent of* ZF *and (ii) it is not known whether or not φ is true.*

(The qualifier 'natural' is necessary in the statement of both of the above facts since one can use Gödelian methods to cook up artificial counterexamples.)

Unfortunately, this is as far as ZF can go: We know that ZF *does* have serious limitations with regard to Σ^1_3-arithmetic. For example, the statement SPHERE is not settled by ZF and this limitation of ZF is not benign. Thus, our situation is this: There are no known serious limitations of ZF with regard to Σ^1_2-arithmetic but there are serious limitations with regard to Σ^1_3-arithmetic. The goal is to obtain a theory which does for Σ^1_3-arithmetic what ZF does for Σ^1_2-arithmetic.

NEED. *We need new axioms A such that there are no known natural serious limitations of* ZF $+ A$ *with regard to Σ^1_3-sentences.*

Roughly, this is the minimal hurdle. The trouble is that it is couched in epistemic and vague terms. But we want to actually prove things about the successes and failures of proposed extensions of ZF. And for this we need a precise, mathematical formulation of the minimal hurdle. To arrive at the correct formulation it will helpful to analyze the success of ZF with regard to Σ^1_2.

2.2.3 Absoluteness. FACT 4 is an epistemic fact. But there is a deep mathematical fact underlying it. It is not just that people have not been smart enough (using the present techniques) to find a natural Σ^1_2 sentence φ such that ZF $+ \varphi \leq$ ZF and φ is known to be independent of ZF; rather, it is *impossible* to find such a statement via set forcing. This is made precise in the following theorem.

THEOREM. (Shoenfield) *Suppose* $V \models \mathrm{ZF}$. *Then*

$$V \models \varphi \quad \leftrightarrow \quad V^{\mathbb{P}} \models \varphi$$

for any partial order $\mathbb{P}$ *and for any* $\underset{\sim}{\Sigma}^1_2$*-sentence* φ.

This theorem says that if you are working in a universe V and you manage to show that a $\underset{\sim}{\Sigma}^1_2$ sentence is *consistent* by constructing a forcing extension $V^{\mathbb{P}}$ then that sentence is actually *true* in V. In this sense, ZF ensures that *consistency implies truth*. As a consequence, one *cannot* use the method of forcing to show that *any* $\underset{\sim}{\Sigma}^1_2$ sentence is independent of ZF. In such a case we say 'ZF freezes $\underset{\sim}{\Sigma}^1_2$'.

2.2.4 The Minimal Hurdle. We are now in a position to state in precise mathematical terms the minimal hurdle that we are setting up for extensions of ZF.

MINIMAL HURDLE. *Freeze* $\underset{\sim}{\Sigma}^1_3$.

Let me underscore three points: (1) The hurdle is *minimal* in that if $\underset{\sim}{\Sigma}^1_3$ is *not* frozen then we *know* that there are serious forms of independence at the third level. (2) The hurdle is *surmountable* in that it is in principle possible to overcome it (as we shall see at the beginning of the next section). (3) The hurdle is *significant* in that the following holds for the theories we shall consider: Any such theory which overcomes the hurdle (a) proves SPHERE and, more importantly, (b) lifts *all* of the results of the early analysts (and much more) to the next level; in particular, any such theory proves $\underset{\sim}{\Pi}^1_1$-determinacy.

Chapter 3

Reflection Principles

Our goal is to find axioms that are justified and overcome the Minimal Hurdle. It turns out that there are many axioms which overcome the Minimal Hurdle; for example, forcing axioms, large cardinal axioms, and axioms of definable determinacy. The issue is whether such axioms are *justified.* There is a standard approach to justifying new axioms, namely, to use *reflection principles* to justify *large cardinal axioms*, axioms that assert the existence of *large* sets or levels V_α. In this chapter I will chart the limitations of this approach.

§3.1 continues the investigation of reflection principles by examining higher order reflection wish second order parameters. §3.2 shows that higher order reflection with third order parameters is inconsistent and discusses two possible ways of skirting inconsistency. §3.3 introduces the reflection principles of Tait. §3.4 establishes limitative results on reflection. I show that Tait's reflection principles cannot overcome the Minimal Hurdle. I also prove the consistency of a broad class of reflection principles and show that these principles (the broadest class of reflection principles known to be consistent relative to large cardinals) cannot overcome the Minimal Hurdle and so cannot yield a significant reduction in incompleteness.

3.1 The Program for Reflection Principles

3.1.1 Indescribables. Up until now we have concentrated on reflection principles of first order with parameters of second order, that is, we have been dealing with $\mathscr{L}_{1,2}$-reflection. And we have seen in §1.3 that such principles provide a uniform justification of the standard system of set theory ZF and certain weak large cardinal axioms such those asserting the existence of inaccessible cardinals. We now turn to richer languages and use stronger reflection principles to derive even stronger large cardinal axioms.

DEFINITION 5. Suppose $C \subseteq \kappa$. C is *closed* iff it contains all of its limit points, that is, points α such that $\alpha = \sup(C \cap \alpha)$. C is *club* in κ iff it is closed and unbounded in κ. Suppose $S \subseteq \kappa$. S is *stationary* in κ iff $S \cap C \neq \varnothing$ for each C which is club in κ. κ is *Mahlo* iff the set of inaccessible cardinals below κ is stationary in κ.

Work in $\mathscr{L}_{2,2}$. We have already seen that Ω is inaccessible and that there are arbitrarily large inaccessibles below it. Since Ω is a limit of inaccessibles there must—by reflection—be a limit of inaccessibles below it. Even more, there must be stationarily many inaccessibles below it: For let C be club in Ω. Since $V \models$'C is a club which is unbounded in the inaccessible Ω' there must by reflection be an α such that $V_\alpha \models$'$C \cap \alpha$ is a club which is unbounded in the inaccessible $\Omega \cap \alpha$'. Now since C is club the inaccessible $\Omega \cap \alpha$ is in C. In other words, C has non-trivial intersection with the class of inaccessible, which is to say that the class of inaccessibles is stationary. Thus Ω is Mahlo. We can now apply reflection to get a Mahlo cardinal and then a proper class of Mahlo cardinals and so on.

To transcend the larger cardinals obtained above we need to enrich our language. The simplest way to define the relevant large cardinals is directly in terms of reflection.

DEFINITION 6. Let Γ be a set of formulas. An infinite cardinal κ is Γ-*indescribable* iff

$$V_\kappa \models \Gamma\text{-reflection}.$$

We can work our way through $\mathscr{L}_{2,2}$ starting with the simplest statements, namely, those which are $\Pi^1_{1,2}$. For example, Spec(MST + $\Pi^1_{1,2}$-reflection) = $\Pi^1_{1,2}$-Indescribables,

also known as Weakly Compacts. Then we can reflect on this to get unboundedly many Weakly Compacts, and the various orders of weakly compacts. The next qualitative jump comes with $\Pi^1_{2,2}$-indescribables. And the path continues with $\Pi^1_{2,2}$-indescribables, $\Pi^1_{3,2}$-indescribables, and so on, until we have to jump up to $\mathscr{L}_{3,2}$ and then $\mathscr{L}_{4,2}$ and then $\cdots$ $\mathscr{L}_{\beta,2}$ $\cdots$ and so on. Doing so yields indescribable cardinals of higher and higher order. Trivially, we have the following:

FACT 7. *$\mathscr{L}_{\beta,2}$-reflection implies the existence of Mahlo, weakly compact, and indescribable cardinals of high-order.*

One cannot squeeze much more out of $\mathscr{L}_{\beta,2}$-reflection. For example, $\mathscr{L}_{\beta,2}$ does not imply the existence of unfoldable or remarkable cardinals. Furthermore, can show that $\mathscr{L}_{\beta,2}$ cannot overcome the Minimal Hurdle—the reasons will become apparent in §3.4. Fortunately, we have not exhausted the resources available to us: we have restricted ourselves to languages involving parameters of only *second* order. We will see that difficulties arise already for parameters of *third* order, that is, in the move to $\mathscr{L}_{1,3}$.

3.1.2 The Program. Given the initial success of reflection principles in yielding large cardinal axioms one might conjecture that *all* large cardinal axioms are ultimately implied by reflection principles. This has been a common view. Thus consider the following:

(1) Some have expressed hope that reflection principles imply all large cardinal axioms. For example, Gödel writes:

> Generally I believe that, in the last analysis, every axiom of infinity should be derivable from the (extremely plausible) principle that V is indefinable, where definability is to be take in [a] more and more generalized and idealized sense. (Wang (1977), p. 325; Wang (1996), p. 285)

Since the most natural way to assert that V is undefinable is via reflection principles and since to assert this in a "more and more generalized and idealized

sense" is to move to languages of higher-order with higher-order parameters, Gödel's proposal amounts to the claim that higher-order reflection principles imply all large cardinal axioms.

(2) Some have claimed that this hope has already been realized. For instance, Martin and Steel write:

> We know of one proper extension of ZFC which is as well justified as ZFC itself, namely ZFC + 'ZFC is consistent'. Extrapolating wildly, we are led to *strong reflection principles*, also known as *large cardinal axioms* (One can fill in some intermediate steps.) These principles assert that certain properties of the universe V of all sets are shared by, or "reflect to", initial segments V_α of the cumulative hierarchy of sets. (Martin & Steel (1989), p. 72)

(3) Some have tried to realize this hope. For example, Tait proved an impressive series of results in this direction in his (1990), (1998a), and (1998b).

So the following program has substantial support.

PROGRAM. (Reflection Principles) *Show that reflection principles imply the standard large cardinal axioms.*

If successful this program would entail that reflection principles clear the Minimal Hurdle by a long shot. A far less ambitious program is the following:

WEAK PROGRAM. (Reflection Principles) *Show that reflection principles overcome the Minimal Hurdle.*

3.1.3 Limitations for the Actualist. Our immediate concern is whether the Weak Program can succeed, that is, whether reflection principles can overcome the Minimal Hurdle. The actualist trips up right at the start. The trouble is that on the actualist view the resources of class quantification are limited. Classes must be treated as *virtual* in the sense of Quine (1969), p. 16–19 and Quine (1986) p. 71–72,

that is, classes must be treated in a manner such that they can be defined away; for if V consists of *all* of the sets then the power*set* of V does not exist. The actualist can arguably handle schematic reflection principles and thereby justify the existence of inaccessible cardinals. But the actualist cannot simulate the full powerset of V and hence cannot simulate full second-order reflection. It is straightforward, for example, to see that (even with the help of iterated truth theories) the actualist cannot simulate enough quantification over classes to ensure that the ordinal height of the universe is weakly compact.

Assume that Ω is weakly compact and work in V, where Ω is also weakly compact. Suppose the actualist lives in $V = V_\Omega$. The actualist can *simulate* talk of classes by iterating the satisfaction predicate (thereby simulating the construction of L beyond V) along any well-ordering definable over L_Ω. For example, one can simulate $L_\Omega(V)$, $L_{\Omega+\Omega}(V)$, and so on. Let ζ_0 be the supremum of the well-orderings definable over $L_\Omega(V)$. Now bootstrap, letting ζ_1 be the supremum of the well-orderings definable over $L_{\zeta_0}(V)$, etc. Finally, let $\zeta = \sup_{n<\omega} \zeta_n$. Clearly, ζ is less than Ω^+. Furthermore, *any* bootstrapping technique will close off at an ordinal $\zeta < \Omega^+$. Now since Ω is weakly compact we have the following: for any transitive M such that $V_\Omega \subseteq M$ and $\mathrm{Card}(M) \leq \Omega$ there is an elementary embedding $j : M \to N$ where N is transitive and $\mathrm{crit}(j) = \Omega$. Choose ζ' such that $\zeta < \zeta' < \Omega^+$, take $M = L_{\zeta'}(V)$, and let $j : M \to N$ be the embedding. With this embedding we can reflect more second-order reflection than the actualist can express, thus obtaining a level V_α below V_Ω which satisfies all of the reflection that can be viewed from the point of view of V_Ω. Therefore the actualist living in V_Ω cannot possibly simulate enough class-quantification to ensure that Ω is weakly compact—the actualist could be in V_α but α is not weakly compact.[5]

The potentialist, in contrast, has no trouble with higher order quantification since for the potentialist talk of V is to be construed as talk of some permissible candidate V_κ and here second order quantification makes sense: the second order quantifiers '$\forall X$' and '$\exists X$' range over $\mathscr{P}(V_\kappa)$, which is just $V_{\kappa+1}$. Likewise higher order quantification makes sense: The β^{th}-order quantifiers '$\forall X^{(\beta)}$' and '$\exists X^{(\beta)}$' range over $\mathscr{P}^\beta(V_\kappa)$, which is just $V_{\kappa+\beta}$. In what follows I will use the term 'class' in a relative fashion. For

example, the phrase 'X is a class' will mean $X \in \mathscr{P}(V_\kappa)$ when we are discussing the model V_κ and it will mean $X \in \mathscr{P}(V_{\kappa+1})$ when we are discussing the model $V_{\kappa+1}$. Strictly speaking, one should say 'X is a class with respect to M' but since the relevant model will be apparent from context no confusion will arise if we drop explicit reference to it.

To summarize: The actualist cannot simulate enough class quantification to state higher-order reflection principles and so the actualist can barely get off the ground, let alone use reflection principles to overcome the Minimal Hurdle. But the potentialist, by treating talk of V as talk of a permissible candidate V_κ, has access to *full* higher order quantification and so *can* formulate higher order reflection principles. For this reason we shall henceforth consider the potentialist point of view. Our question, then, is whether the potentialist can overcome the Minimal Hurdle with reflection principles.

3.1.4 The Potentialist View. The potentialist and the actualist seem to face complementary problems: The actualist can motivate reflection by speaking of the true universe V but has trouble in formulating higher order reflection. The potentialist has no trouble in formulating higher order reflection but has trouble in motivating reflection because according to the potentialist the universe of sets is not determinate. How then does the potentialist speak of V?

Recall our discussion of potentialism in §1.1.5 and §1.3.1. In §1.1.5 I argued that the concept of set involves closure principles. In §1.3.1 I motivated reflection principles by noting that as soon as we fix some precise characterisation of the universe of sets in terms of a closure principle C we see that it falls short, that is, we see that the concept of C-set falls short of the concept of set. Recall also that in that section I introduced for each set of closure principles C the notion of the spectrum $\Lambda(C)$ of C and the notion of a permissible candidate of C. As we strengthen the closure principles we thin out the spectrum.

I suggest that the potentialist regard talk of V as talk of any point in the limit of the above thinning process. The spectrum of the concept of set consists of the

legitimate candidates for the concept of set. The candidates differ from the candidates for the concept of C-set in two respects. First, we can specify the concept of C-set in terms of closure principles but the concept of set cannot be so specified—this is the gist of our above motivation for reflection principles. Second, we can transcend the concept of C-set in the sense that there is a concept of C'-set where the closure principles C' *transcend* the closure principles C. In contrast the concept of set cannot be transcended. It is important to note here the difference between (i) *surpassing* a legitimate candidate V_κ for the concept of set by moving to a higher level, say $V_{\kappa+1}$, and (ii) *transcending* a legitimate candidate V_κ for the concept of set by arriving at a concept of set with essentially richer closure principles. The potentialist as I am understanding him believes that a legitimate candidate for the concept of set can be surpassed but not transcended. The potentialist believes that when one moves beyond a legitimate candidate it is just more of the same, one never reaches a level with essentially richer closure conditions.

There is an obvious objection to this last point. Suppose that V_κ and V_λ are legitimate candidates for the concept of set where $\kappa < \lambda$. Then it would seem that V_λ satisfies stronger closure principles since it has a level V_κ which satisfies all of the closure principles inherent in the concept of set. The potentialist reply is that although this may hold true of the concept of C-set the situation with the concept of set is quite different. In the case of the concept of C-set the closure principles can be described and so we can refer to the first element V_κ of $\Lambda(C)$ within the second element V_κ of $\Lambda(C)$ and so it is indeed true that V_λ satisfies stronger closure principles than V_κ. But in the case of the full concept of set the second legitimate candidate does not satisfy that it has a level which meets the concept of set. This is because the closure principles of the concept of set cannot be specified.

On this version of the potentialist view the legitimate candidates for the concept of set are *order indiscernibles*. V is any one of a maximally closed level. No legitimate candidate can be defined and apart from their order they cannot be distinguished one from the other. It is in this way that the potentialist preserves talk of V while maintaining that any set can be surpassed.

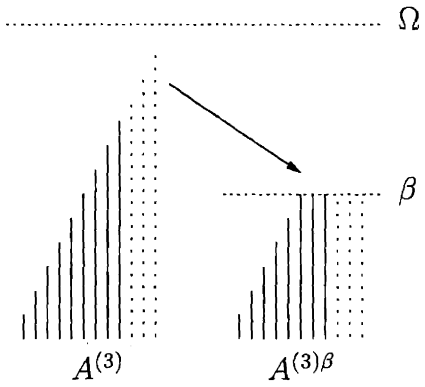
3.2 Higher-Order Parameters

Having thus exhausted the resources of $\mathscr{L}_{\beta,2}$-reflection let us allow parameters of third and higher order. The first issue to be dealt with is how to relativise such parameters.

3.2.1 Relativisation. Recall that the relativisation $X^{(2)\alpha}$ of a second order class $X^{(2)}$ to the level V_α is $X^{(2)} \cap V_\alpha$. This is how someone living in V_α views $X^{(2)}$. Now consider a third order class $Y^{(3)}$. It consists of various second order classes $X^{(2)}$. The question is: how we are to relativise $Y^{(3)}$ to V_α; how does someone living in V_α view $Y^{(3)}$? As a start we might ask how the members $X^{(2)}$ of $Y^{(3)}$ are viewed. Well, we already dealt with this question: each member $X^{(2)}$ is viewed as $X^{(2)\alpha}$. So since $Y^{(3)}$ consists of members of the form $X^{(2)}$ it is natural to take the reflected version of $Y^{(3)}$ to consist of the reflections of the members, that is, $Y^{(3)\alpha} = \{X^{(2)\alpha} \mid X^{(2)} \in Y^{(3)}\}$. More generally, the relativisation $Y^{(\beta+1)\alpha}$ of $Y^{(\beta+1)}$ to V_α is defined inductively by: $Y^{(\beta+1)\alpha} = \{X^{(\beta)\alpha} \mid X^{(\beta)} \in Y^{(\beta+1)}\}$. And similarly for β a limit. This is the most natural course to take. I shall call this version of relativisation *unconstrained relativisation.*

3.2.2 Inconsistency. Unfortunately, reflection on third-order parameters with unconstrained relativisation is inconsistent. Consider the class third-order class $A^{(3)}$ consisting of all second-order classes that consist of initial segments of the ordinals that are *bounded.* Thus $A^{(3)}$ contains the class of ordinals up to 1, and the class of ordinals up to 2 and, more generally, the class of ordinals up to α for every ordinal α. More precisely, letting $[0 \cdots \alpha)$ denote the second-order class of ordinals less than β we have $A^{(3)} = \big[\,[0 \cdots \alpha) \mid \alpha \in \Omega\,\big]$.

Now look at what happens when we relativise $A^{(3)}$ to a lower level V_β. As described above, the relativisation of $A^{(3)}$ to V_β is achieved by cutting each element $X^{(2)}$ of $A^{(3)}$ off at V_β. Now take an element of $A^{(3)}$ which is *taller* than β. When we relativise this element to V_β it gets cut back to the class of *all ordinals below* α. Here is the picture:



So it is true in V that each member of $A^{(3)}$ is bounded in the ordinals, but this is no longer true of its relativisation $A^{(3)\beta}$, since for $\alpha > \beta$ the class $[0 \cdots \alpha)$, which is bounded in V, gets relativised to the class $[0 \cdots \beta)$, which is unbounded in V_β (all of the lines on the left hand side of the diagram are bounded below the height Ω of the universe V but many of the lines on the right hand side of the diagram are unbounded in the height β of the universe V_β). So we have proved:

FACT 8. *$\mathscr{L}_{1,3}$-reflection is inconsistent.*

3.2.3 Two Fixes. There are two possible ways to save reflection principles:

(1) Constrain the language.

(2) Constrain the form of relativisation.

Let us consider each approach in turn.

3.3 Tait

The first fix is due to Tait. Notice that the above counterexample arises because we are allowing ourselves to make statements about what is *not* in a third- or higher-order class: all facts about what *is* in a third- or higher-order class relativise down correctly. So we might avoid inconsistency by refraining from making "negative" statements. This is true but why should it matter? It is hard to see any motivation,

other than the desire to skirt inconsistency, for restricting the language. But doing so leads to interesting results of Tait.

3.3.1 Γ_n-reflection. Our first task is to give a precise description of the restrictions to be imposed on the language. We will then be in a position to introduce both Tait's notion of Γ_n-reflection and related notions that will be of use later.

DEFINITION 9. Work in $\mathscr{L}_{<\omega}$. Let lower case letters $x, y, z, \ldots$ range over the first type and upper case letters $X, Y, Z, \ldots$ range over higher types. A formula $\varphi \in \mathscr{L}_{<\omega}$ is *first-order* iff all of the bound variables of φ are of the first type. (Notice that we allow free parameters of higher order). We *relativise* objects X and formulas $\varphi \in \mathscr{L}_{<\omega}$ to V_β in the manner described above. (For X of first order $X^\beta = X$; for X of higher order (i.e. second or greater) and of type $(\tau_1 \ldots \tau_n)$, $X^\beta = \{(X_1^\beta \ldots X_n^\beta \mid (X_1 \ldots X_n) \in X\}$; for $\varphi(X_1 \ldots X_n) \in \mathscr{L}_{FT}$, $\varphi^\beta(X_1^\beta \ldots X_n^\beta)$ is the result of restricting the bound variables of first type to $(V_\beta, ()) \equiv V_\beta$ and those of type $\tau = (\tau_1 \ldots \tau_n)$ to $\mathscr{P}((V_\beta, \tau_1) \times \cdots \times \mathscr{P}(V_\beta, \tau_N))$.)

(1) $\varphi \in \mathscr{L}_{<\omega}$ is *positive* ($\varphi \in \mathscr{L}^P_{<\omega}$) if φ is the minimal class of formulas containing $x = y$, $x \neq y$, $x \in y$, $x \notin y$, $x \in Y$, $x \notin Y$, $X = Y$, $(X_1 \ldots X_n) \in Y$ and closed under $\wedge$, $\vee$, $\forall$ and $\exists$.

(2) Γ_n is the pointclass

$$\{\varphi \in \mathscr{L}^P_{<\omega} \mid \varphi = \forall X_1 \exists Y_1 \cdots \forall X_n \exists Y_n\ \psi \text{ where}$$
$$\text{i) } \psi \text{ is first-order,}$$
$$\text{ii) } X_i \text{ is second-order, and}$$
$$\text{iii) } Y_i \text{ is of arbitrary finite type}\}.$$

(3) Suppose $X \subseteq \kappa$. X is *n-reflective in* κ iff

$$\forall X_1 \cdots \forall X_m\, [V_\kappa \models \varphi(X_1 \cdots X_n) \rightarrow \exists \beta \in X\, (\varphi^\beta(X_1^\beta \cdots X_n^\beta))]$$

for all $\varphi(X_1 \cdots X_n) \in \Gamma_n$ with X_i of arbitrary order. Γ_n-*reflection* is the schema stating Ω is n-reflective in Ω.

(4) Let $X \subseteq \kappa$ and $0 < n < \omega$. Let the variables $\mathscr{T}$ with or without subscripts range over sequences $\langle T_\alpha \mid \alpha < \kappa \rangle$ such that $T_\alpha \subseteq V_\alpha$ and for $T \subseteq V_\kappa$ let $G_{T,\mathscr{T}} = \{\alpha < \kappa \mid T \cap V_\alpha = \mathscr{T}(\alpha)\}$ be the set of points where T *correctly guesses* $\mathscr{T}$. IN^κ_γ is defined recursively as follows:

$$\begin{aligned}
\text{IN}^\kappa_0 &= \{X \subseteq \kappa \mid X \text{ is stationary in } \kappa\} \\
\text{IN}^\kappa_{\gamma+1} &= \{X \subseteq \kappa \mid \forall \mathscr{T} \exists T\, (G_{T,\mathscr{T}} \cap X \in \text{IN}^\kappa_\gamma)\} \text{ provided } \text{IN}^\kappa_\gamma \neq \varnothing \\
\text{IN}^\kappa_\lambda &= \bigcap\nolimits_{\gamma<\kappa} \text{IN}^\kappa_\gamma, \text{ for } \lambda \text{ limit.}
\end{aligned}$$

3.3.2 Ineffables. As we shall see Γ_n-reflection has significant strength. To measure this strength we introduce the appropriate large cardinal notions, namely, *ineffable* cardinals (introduced by Kunen and Jensen) and their generalisations (introduced by Baumgartner).

DEFINITION 10. Let the variable κ range over infinite cardinals. Let $X \subseteq \kappa$ and $0 < n < \omega$. Let the variables $\mathscr{S}$ with or without subscripts range over sequences $\langle S_\alpha \mid \alpha < \kappa \rangle$ such that $S_\alpha \subseteq \alpha$ and for $S \subseteq \kappa$ let $G_{S,\mathscr{S}} = \{\alpha < \kappa \mid S \cap \alpha = \mathscr{S}(\alpha)\}$ be the set of points where S *correctly guesses* $\mathscr{S}$. Similarly, let the variables $\mathscr{S}^n$ with or without subscripts range over sequences $\langle S_{\alpha_1 \ldots \alpha_n} \mid \alpha_1 < \cdots \alpha_n < \kappa \rangle$ such that $S_{\alpha_1 \ldots \alpha_n} \subseteq \alpha_1$ and for $S \subseteq \kappa$ let $G_{S,\mathscr{S}^n} = \{(\alpha_1 \ldots \alpha_n) \in [\kappa]^n \mid S \cap \alpha_1 = \mathscr{S}^n(\alpha_1 \ldots \alpha_n)\}$ be the set of points where S *correctly guesses* $\mathscr{S}^n$.

(1) X is *ineffable in* κ iff $\forall \mathscr{S} \exists S\, (G_{S,\mathscr{S}} \cap X$ is stationary in $\kappa)$.

(2) X is *n-ineffable in* κ iff $\forall \mathscr{S}^n \exists S \exists W\, (W \subseteq X \wedge [W]^n_< \subseteq G_{S,\mathscr{S}^n} \wedge W$ is stationary in $\kappa)$, where $[W]^n_<$ is the set of strictly increasing n-sequences of ordinals from W.

(3) $\mathrm{In}^{\kappa}_{\gamma}$ is defined recursively as follows:

$$\mathrm{In}^{\kappa}_{0} = \{X \subseteq \kappa \mid X \text{ is stationary in } \kappa\}$$
$$\mathrm{In}^{\kappa}_{\gamma+1} = \{X \subseteq \kappa \mid \forall \mathscr{S} \exists S\, (G_{S,\mathscr{S}} \cap X \in \mathrm{In}^{\kappa}_{\gamma})\} \text{ provided } \mathrm{In}^{\kappa}_{\gamma} \neq \varnothing$$
$$\mathrm{In}^{\kappa}_{\lambda} = \bigcap\nolimits_{\gamma<\kappa} \mathrm{In}^{\kappa}_{\gamma}, \text{ for } \lambda \text{ limit.}$$

κ is *completely ineffable* iff $\forall \gamma < (2^{\kappa})^{+}\, (\mathrm{In}^{\kappa}_{\gamma} \neq \varnothing)$.

3.3.3 Connection. We have the following connections:

THEOREM 11. (Tait) *Let $X \subseteq \kappa$. The following are equivalent*

(1) *X is n-reflective*

(2) $X \in \mathrm{IN}^{\kappa}_{n}$.

LEMMA 12. (Baumgartner, Tait) *Let $X \subseteq \kappa$. If*

(1) $X \in \mathrm{IN}^{\kappa}_{n}$ *then*

(2) *X is n-ineffable.*

COROLLARY 13. *Suppose X is n-reflective. Then X is n-ineffable.*

The central result is the following:

THEOREM 14. (Tait) *Suppose V_{κ} satisfies Γ_n-reflection. Then κ is n-ineffable.*

Tait's theorem shows that Γ_n-reflection has great strength. The key question is whether it has enough strength to overcome the Minimal Hurdle. The above theorem does not settle this question; n-ineffable cardinals are too weak. But it does provide hope; perhaps this is just the beginning and Γ_n-reflection can yield much more.

3.3.4 Consistency. The first issue we must face is whether Γ_n-reflection is consistent. This may seem surprising since up until now reflection principles have played a foundational role. For we derived first and second order reflection principles from

our analysis of the concept of set and then we used reflection principles to justify ZFC and much more. The situation with Γ_n-reflection is, however, different. In contrast to the earlier reflection principles it is difficult to maintain that it follows from an analysis of the concept of set. For we have had to impose an ad hoc restriction on the language in order to skirt inconsistency. Fortunately, large cardinals imply the consistency of Γ_n-reflection.

THEOREM 15. (Tait) *Suppose that κ is a measurable cardinal. Then V_κ satisfies Γ_n-reflection for all n.*

Thus the strength of Γ_n-reflection is somewhere between that of a cardinal which is n-ineffable for all n, and a measurable cardinal. If this upper bound were anywhere near optimal then Γ_n-reflection would overcome the Minimal Hurdle. This question was open.

3.4 Limitations

3.4.1 The First Fix. In this section I will show that Γ_n-reflection cannot overcome the Minimal Hurdle.

DEFINITION 16. For $\alpha \geq \omega$ the *Erdös cardinal* $\kappa(\alpha)$ is the least κ such that $\kappa \to (\alpha)^{<\omega}_2$, i.e. such that for each partition $P : [\kappa]^{<\omega} \to 2$ there is an $X \in [\kappa]^\alpha$ such that $\mathrm{Card}(P[[X]^n]) = 1$ for all $n < \omega$, where $P[Y] = \{P(a) \mid a \in Y\}$.

THEOREM 17. *Assume $\kappa = \kappa(\omega)$ exists. Then there is an $\delta < \kappa$ such that V_δ satisfies Γ_n-reflection for all n.*

Proof. We need a preliminary lemma.

LEMMA. (Silver) *Assume $\alpha \geq \omega$. Then the following are equivalent*

(1) $\kappa \to (\alpha)^{<\omega}_2$

(2) *For all structures $\mathscr{M}$ such that*

(a) $\mathrm{Card}(\mathscr{L}(\mathscr{M})) = \omega$ *and*

(b) $\kappa \subseteq |\mathscr{M}|$

there is an $X \in [\kappa]^\alpha$ *which is a set of indiscernibles for* $\mathscr{M}$.

Proof. (1) $\rightarrow$ (2): For each partition $P : [\kappa]^{<\omega} \rightarrow 2$ we have a homogeneous set X—that is, an $X \in [\kappa]^\alpha$ such that $\mathrm{Card}(P[X]^n) = 1$ for all $n < \omega$. Take $P : [\kappa]^{<\omega} \rightarrow 2$ where

$$P(\zeta_1 \cdots \zeta_n) = \begin{cases} 0 & \text{if } \mathscr{M} \models \varphi_n[\zeta_1 \cdots \zeta_{\mathrm{fv}(n)}] \\ 1 & \text{otherwise} \end{cases}$$

where $\{\varphi_n\}$ enumerates $\mathscr{L}(\mathscr{M})$ and $\mathrm{fv}(n)$ is the number of free variables in φ_n.

(2) $\rightarrow$ (1): Fix $P : [\kappa]^{<\omega} \rightarrow 2$. Consider $\mathscr{M} = (\kappa, \in, P|[\kappa]^1 \cdots P|[\kappa]^n \cdots)$. Let $X \in [\kappa]^\alpha$ be a set of indiscernibles for $\mathscr{M}$. Then X is homogeneous for P since for $(\zeta_1 \cdots \zeta_n)$ and $(\zeta_1' \cdots \zeta_n')$, both in $[X]^n$, we have $P|[\kappa]^n(\zeta_1 \cdots \zeta_n) = P|[\kappa]^n(\zeta_1' \cdots \zeta_n')$ by indiscernibility. □

Following the proof of Silver-Reinhardt that there is a totally indescribable cardinal less than κ we consider the structure $\mathscr{N} = (V_\kappa, \in, <)$ where $<$ is a well-ordering of V_κ. Let $I' = \{\iota_k'\}$ be the indiscernibles of $\mathscr{N}$ given by Silver's theorem. Take the Skolem hull of these indiscernibles and collapse, letting

$$\mathscr{M}' = \mathrm{Hull}^{\mathscr{N}}(I'),$$
$$\pi : \mathscr{M}' \cong \mathscr{M} = \mathrm{collapse}(\mathscr{M}), \text{ and}$$
$$I = \pi(I').$$

Notice that I is a set of indiscernibles for $\mathscr{M}$ and that by including the well-ordering $<$ in $\mathscr{N}$ we ensure that these indiscernibles, which we will enumerate as $\{\iota_k\}$, obey the key properties (with respect to $\mathscr{M}$) obeyed by the Silver indiscernibles (with respect to L). See Kanamori (1997, Ch. 9) for details.

Now let $h : I \rightarrow I$ be order preserving and such that $\mathrm{crit}(h) = \iota_k$. This map uniquely extends to an elementary embedding $j : \mathscr{M} \rightarrow \mathscr{M}$ with $\mathrm{crit}(j) = \iota_k$. We aim to show that $V_{\iota_k}^M$ satisfies Γ_n-reflection for all n. This statement is to be regarded as a statement of M, through the coding of classes as sets.

Work in M. Fix $\varphi(A_1 \cdots A_m) \in \Gamma_n$. Assume

$$V_{\iota_k}^M \models \varphi(A_1 \cdots A_m).$$

We would like to show that

$$V_{j(\iota_k)}^M \models \exists \alpha < j(\iota_k)\ \varphi^\alpha(j(A_1)^\alpha \cdots j(A_m)^\alpha)$$

since this would imply

$$V_{\iota_k}^M \models \exists \alpha < \iota_k\ \varphi^\alpha(A_1^\alpha \cdots A_m^\alpha).$$

We have that

$$V_{j(\iota_k)}^M \models \varphi^{\iota_k}(A_1 \cdots A_m).$$

So we would be done if $j(A)^{\iota_k} = A$. Unfortunately, this is not always true. For example, consider

$$A^{(3)} = \big[[0 \cdots \alpha) \mid \alpha < \iota_k\big]$$

and notice that $j(A^{(3)})^{\iota_k} \neq A^{(3)}$ since the former picks up $[0 \cdots \iota_k)$.

Notice, however, that it suffices to prove the following lemma where, for notational convenience, we let $A^* = j(A)^{\iota_k}$.

LEMMA. *Suppose* $\varphi(A_1 \cdots A_m) \in \mathscr{L}_\beta^P$ *where* $\beta < \iota_n$. *Then if*

$$V_{\iota_k}^M \models \varphi(A_1 \cdots A_m)$$

then

$$V_{\iota_k}^M \models \varphi(A_1^* \cdots A_m^*).$$

Proof. Here the second φ is really the natural translation of the first—occurrences of '$\in$' between class parameters are reinterpreted in the shift to sets and the class quantifiers become set quantifiers. The proof is by induction on n.

BASE CASE: Assume $\varphi(A_1 \cdots A_m) \in \Gamma_0$, i.e. $\varphi(A_1 \cdots A_m)$ is a first order positive formula with parameters of arbitrary type. It is easy that $\varphi(A_1 \cdots A_m)$ reflects

to the club C of points below ι_k which are closed under the Skolem functions for $\varphi(A_1 \cdots A_m)$. Thus, $\varphi(j(A_1) \cdots j(A_m))$ reflects to the club of points $j(C)$ below $j(\iota_k)$, by elementarity. But $j(C) \cap C = C$ and, since C is unbounded in ι_k and $j(C)$ is club, this implies that $\iota_k \in j(C)$, that is, $\varphi(j(A_1) \cdots j(A_m))$ reflects to ι_k.

INDUCTION STEP: Assume the claim is true for $\psi \in \Gamma_n$. The following are equivalent:

$$
\begin{aligned}
&V^M_{\iota_k} \models \forall X \exists Y \psi(X, Y, \vec{A}) \\
&V^M_{\iota_k} \models \exists Y \psi(C, Y, \vec{A}), \text{ for chosen } C \\
&V^M_{\iota_k} \models \psi(C, B, \vec{A}) \text{ some } B \\
&V^M_{\iota_k} \models \psi(C^*, B^*, \vec{A}^*) \\
&V^M_{\iota_k} \models \exists Y \psi(C^*, Y, \vec{A}^*) \text{ chosen } C, \text{ which was arbitrary} \\
&V^M_{\iota_k} \models \forall X \exists Y \psi(X, Y, \vec{A}^*).
\end{aligned}
$$

The fourth line is equivalent to the third by the induction hypothesis. The key point is that the fifth line is equivalent to the fourth since $C = C^*$, which is true because the universal quantifier $\forall X$ ranges over *second-order* classes.

This completes the proof of the lemma. □

Now applying π^{-1} we have that

$$\mathscr{M}' \models \big(V_{\pi^{-1}(\iota_k)} \models \Gamma_n\text{-reflection, for all } n\big)$$

and hence that

$$V_\kappa \models \big(V_{\pi^{-1}(\iota_k)} \models \Gamma_n\text{-reflection, for all } n\big).$$

□

Notice that in the proof we made key use of the fact that the universal quantifiers are second-order. Let $\mathscr{L}^{P_2}_\beta$ be $\mathscr{L}^P_\beta$, except where all universal quantifiers are second-order. So, using this terminology, we showed that $\mathscr{L}^{P_2}_{<\omega}$-reflection is consistent and

weaker than an Erdös cardinal. In fact, the proof shows that one can allow *existential* quantifiers of *any* ordinal order. Thus we have

THEOREM 18. *Assume $\kappa = \kappa(\omega)$ exists. Then there is an $\delta < \kappa$ such that V_δ satisfies $\mathscr{L}_\beta^{P_2}$-reflection for all b.*

This is the most powerful reflection principle currently known to be consistent.

QUESTION. *Are higher-order reflection principles with third-order universal quantifiers consistent?*

We are now almost in a position to show that no reflection principle known to be consistent can overcome the Minimal Hurdle. First we need a lemma.

LEMMA 19. (Woodin (1982)) *The following are equivalent:*

(1) *$X^\#$ exists for all X.*

(2) *If G_1 is set generic over V and G_2 is set generic over $V[G_1]$, then $V[G_1]$ and $V[G_2]$ satisfy the same $\underset{\sim}{\Sigma}^1_3$ statements (with parameters from $V[G_1]$).*

Let us say that a set of axioms A is *generically invariant* if and only if ZF + A is preserved under set-size forcing extensions. For example, CH is not generically invariant but ZF and all of the reflection principles that we have considered thus far are generically invariant.

COROLLARY 20. *Suppose* ZF + A. *If* A *is generically invariant and* ZF + A *freezes $\underset{\sim}{\Sigma}^1_3$ then $X^\#$ exists for all X.*

Proof. Assume that A is generically invariant and that ZFC + A freezes $\underset{\sim}{\Sigma}^1_3$. Let G_1 be set generic over V. Since A is generically invariant, $V[G_1] \models$ ZF + A. Suppose G_2 is set generic over $V[G_1]$. Since ZF + A freezes $\underset{\sim}{\Sigma}^1_3$ we have that $V[G_1]$ and $V[G_2]$ satisfy the same $\underset{\sim}{\Sigma}^1_3$ statements (with parameters from $V[G_1]$). Thus, by the lemma we have that $X^\#$ exists for all X. □

Thus if a reflection principle overcomes the Minimal Hurdle it must imply that $X^{\#}$ exists for all X. Assume that $\kappa(\omega)$ exists. Then $L \models$'$\kappa(\omega)$ exists'. It follows from Theorems 17 and 18 that Γ_n-reflection and $\mathscr{L}_\beta^{P_2}$-reflection hold in some L_δ. Since this model does not satisfy that $X^{\#}$ exists for all X we have that neither Γ_n-reflection and $\mathscr{L}_\beta^{P_2}$-reflection can overcome the Minimal Hurdle. And since the latter is the strongest reflection principle currently known to be consistent we have the following:

THEOREM 21. *No reflection principle known to be consistent can overcome the Minimal Hurdle.*

3.4.2 The Second Fix. Let us turn to the second approach, namely, the attempt to save reflection principles by constraining the form of relativisation. Although it is difficult to see *exactly* how to incorporate enough definability constraints to avoid the problems presented by examples like $A^{(3)}$, one can point to a general condition which must be satisfied by any reasonable form of constrained relativisation or even, arguable, by *any* form of relativisation. Let M be a model of set theory and suppose that there is an elementary embedding $j : M \to M$ which is not the identity. Let κ be the least ordinal moved by j, the *critical point* of j. Since j is an elementary embedding *everything* which is true of a set $a \in M$ is also true of $j(a) \in M$, as far as M is concerned. That is, from the point of view of M the sets a and $j(a)$ are indistinguishable. (Of course, we, on the outside, can see the difference since, unlike M, we can see j). Now consider a higher-order class $X^{(\beta)}$ over V_κ^M, the universe of M cut off at the κ^{th} level. When one applies j to this class one gets $j(X^{(\beta)})$, which, from the point of view of M, is indistinguishable from $X^{(\beta)}$. And it seems that, in the very least, the modified form of relativisation should be such that when you relativise $j(X^{(\beta)})$ back down to V_κ^M you get back what you started with, that is, relativisation should be such that $j(X^{(\beta)})^\kappa = X^{(\beta)}$. This is the *definability constraint* on relativisation. The proof of Theorem 17 establishes the following:

THEOREM 22. *No reflection principle which meets the definability constraint can overcome the Minimal Hurdle.*

3.4.3 Conclusion. The standard way of justifying new axioms is in terms of reflection principles. This chapter has been devoted to charting out the limitations of this method. In §3.1.3 we saw that the actualist does not have the resources to formulate strong reflection principles. In contrast, the potentialist can formulate and arguably justify strong reflection principles by making use of the notion of a legitimate candidate for V. Unfortunately, reflection principles which allow third order are inconsistent. There are two fixes. First, one can restrict the language in the manner of Tait. In §3.4.1 I showed that Tait's principle of Γ_n-reflection cannot overcome the Minimal Hurdle. Furthermore, I established the consistency of a series of reflection principles which are stronger than Γ_n-reflection and showed that they cannot overcome the Minimal Hurdle. The current situation is that *no* reflection principle which is known to be consistent can overcome the minimal hurdle. Second, one can leave the language untouched and restrict instead the nature of relativisation. I argued for a natural restriction—the definability constraint—and in §3.4.2 showed that no reflection principle meeting this constraint can overcome the Minimal Hurdle.

In the next Chapter I will introduce a new kind of principle—the extension principles.

3.5 Appendix

In this appendix I will provide an alternative approach to the above theorem. The reason for doing this is twofold. First, this approach yields further information, for instance, that n-reflective cardinals relativise to L. Second, it is hoped that this approach will shed some light on the open question stated above on page 59.

Recall that Tait showed that for $X \subseteq \kappa$, X is n-reflective iff $X \in \text{IN}_n^\kappa$. Above we argued in terms of the property of being n-reflective; here we argue in terms of IN_n^κ. We begin with a few definitions. If $\mathscr{I}$ is an ideal over κ then the *dual filter* is

$$\mathscr{I}^* = \{X \mid \kappa \setminus X \in \mathscr{I}\}$$

and the set of $\mathscr{I}^*$-*stationary sets* is

$$\mathscr{I}^+ = \mathscr{P}(\kappa) \setminus \mathscr{I}.$$

Set

$$\begin{aligned}\mathscr{I} &= \mathscr{P}(\kappa) \setminus \mathrm{In}_0^\kappa \\ \mathscr{I}_\alpha &= \mathscr{P}(\kappa) \setminus \mathrm{In}_\alpha^\kappa \\ \mathscr{I}_{\mathrm{ci}} &= \mathscr{P}(\kappa) \setminus \bigcap_{\alpha<(2^\kappa)^+} \mathrm{In}_\alpha^\kappa\end{aligned}$$

and let $\mathscr{F} = \mathscr{I}^*$, $\mathscr{F}_\alpha = \mathscr{I}_\alpha^*$, and $\mathscr{F}_{\mathrm{ci}} = \mathscr{I}_{\mathrm{ci}}^*$.

PROPOSITION 23. (Baumgartner) *$\mathscr{F}_{\gamma+1}$ and $\mathscr{F}_{\mathrm{ci}}$ are normal filters.*

Proof. Suppose, for contradiction, that $\mathscr{F}_{\gamma+1}$ is *not* normal. Pick $X \in \mathscr{I}_{\gamma+1}^+ = \mathrm{In}_{\gamma+1}^\kappa$ and a regressive function $f : X \to \kappa$ such that $\forall \eta < \kappa \ (f^{-1}[\{\eta\}] \notin \mathrm{In}_{\gamma+1}^\kappa)$. For each $\eta < \kappa$, choose $\mathscr{S}_\eta$ such that for all S_η, $G_{S_\eta,\mathscr{S}_\eta} \cap X \notin \mathrm{In}_\gamma^\kappa$. We seek a sequence $\mathscr{S}$ such that a guess S will serve as a guess for some $\mathscr{S}_\eta$, thus contradicting our choice of $\mathscr{S}_\eta$. Since we cannot fix η beforehand we choose our sequence to encode $\mathscr{S}^{f(\alpha)}(\alpha)$ at the α^{th}-stage and ensure that a guess will "freeze" all $f(\alpha)$ to some fixed η. (This will become clearer as we proceed.) To achieve these two ends simultaneously we code pairs of ordinals via an injective function $\pi : \kappa \times \kappa \to \kappa$ and take as our coding sequence $\mathscr{S}$ such that

$$\mathscr{S}(\alpha) = \pi[\{f(\alpha)\} \times \mathscr{S}_{f(\alpha)}(\alpha)] \cap \alpha.$$

We are interested in

$$C = \{\alpha < \kappa \mid \pi[\alpha \times \alpha] \subseteq \alpha\},$$

the club of closure points of π. By induction we have that $\mathrm{In}_\gamma^\kappa$ contains the club filter (using normality at successor stages) and thus $C \in \mathrm{In}_\gamma^\kappa$. And by the ineffability of X we have a guess S such that

$$G_{S,\mathscr{S}} \cap X \in \mathrm{In}_\gamma^\kappa$$

and hence, by the previous sentence,

$$G_{S,\mathscr{S}} \cap X \cap C \in \mathrm{In}^{\kappa}_{\gamma}.$$

Now notice what our guess S achieves: First, it is easy to see that there must be an η such that $f(\alpha) = \eta$ for all $\alpha \in G_{S,\mathscr{S}} \cap X \cap C$. Second, for $\beta < \gamma$ both in $G_{S,\mathscr{S}} \cap X \cap C$ we have

$$\mathscr{S}_{f(\gamma)}(\gamma) \cap \beta = \mathscr{S}_{f(\beta)}(\beta)$$

which implies (since we have "frozen" all $f(\alpha)$ to η),

$$\mathscr{S}_{\eta}(\gamma) \cap \beta = \mathscr{S}_{\eta}(\beta).$$

Thus, letting S_η be such that

$$S_\eta|\alpha = \begin{cases} \mathscr{S}_\eta(\alpha) & \text{for } \alpha \in G_{S,\mathscr{S}} \cap X \cap C \\ \varnothing & \text{otherwise,} \end{cases}$$

we have a guess S_η such that $G_{S_\eta,\mathscr{S}_\eta} \cap X \cap C$, which contradicts our choice of $\mathscr{S}_\eta$. □

PROPOSITION 24. $\mathrm{In}^{\kappa}_{\gamma} = \mathrm{IN}^{\kappa}_{\gamma}$

Proof. The inclusion $\mathrm{IN}^{\kappa}_{\gamma} \subseteq \mathrm{In}^{\kappa}_{\gamma}$ is immediate. For the other inclusion, we proceed by induction. CASE 0 and CASE λ are immediate. CASE $\gamma + 1$: Assume $X \in \mathrm{In}^{\kappa}_{\gamma+1} \neq \varnothing$. We begin by constructing a function which will translate $\mathscr{T}$-sequences into $\mathscr{S}$-sequences: Since κ is ineffable it is a limit of inaccessibles. Let C be the closure of the set of inaccessibles below κ and let $\pi : \kappa \cong V_\kappa$ be such that for all $\alpha \in C$, $\pi|\alpha : \alpha \cong V_\alpha$. Notice that $C \in \mathrm{In}^{\kappa}_{\gamma}$ since, for every successor $\eta + 1$, normality implies that $\mathrm{In}^{\kappa}_{\eta+1}$ contains the club filter. (Suppose $\mathscr{F}$ is a normal filter on κ and suppose, for contradiction, that C is a club in κ such that $C \notin \mathscr{F}$. Let $f(\eta) = C \cap \eta$. $f|\kappa \setminus (\{C \cup \{0\})$ is regressive and $\kappa \setminus (C \cup \{0\})$ is $\mathscr{F}$-stationary. Thus, by normality, there is an $\alpha < \kappa$ such that $f^{-1}[\alpha]$ is $\mathscr{F}$-stationary, which contradicts the fact that $f^{-1}[\alpha]$ is bounded.)

Now fix $\mathscr{T}$. We seek T such that $G_{T,\mathscr{T}} \cap X \in \text{IN}^\kappa_\gamma$. Let $\mathscr{S} = \pi^{-1}[\mathscr{T}]$, choose S such that $G_{S,\mathscr{S}} \cap X \in \text{In}^\kappa_\gamma$, and let $T = \pi[S]$. Notice that $G_{T,\mathscr{T}} \cap X \cap C = G_{S,\mathscr{S}} \cap X \cap C \in \text{In}^\kappa_\gamma$. (Note that the statement of membership is true even for γ limit). Since In^κ_γ is a filter, we have $G_{T,\mathscr{T}} \cap X \in \text{In}^\kappa_\gamma$ and hence, by our induction hypothesis, $G_{T,\mathscr{T}} \cap X \in \text{IN}^\kappa_\gamma$. $\square$

The key idea in the following lemma is due to Jensen and Kunen.

LEMMA 25. *Suppose* $X \in \text{In}^\kappa_\gamma$ *and* $X \in L$. *Then* $L \models X \in \text{In}^\kappa_\gamma$.

Proof. We show by induction on γ that $\text{In}^\kappa_\gamma \cap L \subseteq (\text{In}^\kappa_\gamma)^L$. CASE 0: $\text{In}^\kappa_0 \cap L \subseteq (\text{In}^\kappa_0)^L$ is immediate since a club in L is a club in V. CASE λ: This is also immediate. CASE $\gamma + 1$: Assume $\text{In}^\kappa_\gamma \cap L \subseteq (\text{In}^\kappa_\gamma)^L$. We must show $\text{In}^\kappa_{\gamma+1} \cap L \subseteq (\text{In}^\kappa_{\gamma+1})^L$—that is, for each $X \in \text{In}^\kappa_{\gamma+1} \cap L$ we must show that $\forall \mathscr{S} \in L \; \exists S \in L \; (G_{S,\mathscr{S}} \cap X) \in (\text{In}^\kappa_\gamma)$. So fix $\mathscr{S} \in L$. We have a witness S in V. We shall show that this witness is in fact in L. To begin, note that $\forall \alpha < \kappa \; (S \cap \alpha \in L)$ since, for cofinally many $\alpha < \kappa$, $S \cap \alpha = \mathscr{S}(\alpha) \in L$. Now we want to argue that

$$(V_\kappa, S) \models \text{“}S \in L\text{”}$$

since

$$\forall \alpha < \kappa \; (V_\alpha, S \cap \alpha) \models \text{“}S \cap \alpha \in L\text{”}$$

and κ is Π^1_1-indescribable. Thus we seek a Σ^1_1 sentence φ such that

$$(V_\kappa, S) \models \varphi(S) \text{ iff } S \in L.$$

The key points are (1) $S \in L$ iff $\exists \alpha < (\kappa^+)^L \; (S \in L_\alpha)$ and (2) each such L_α can be

coded by a subset of L_κ. Thus we take

$$\begin{aligned}\varphi(S) \equiv \exists M, E, \pi, s \text{ such that}\\ &(a)\ M \subseteq V_\kappa,\ E \subseteq V_\kappa \times V_\kappa,\ s \in M,\\ &(b)\ (M, E) \text{ is well-founded, extensional and set-like,}\\ &(c)\ (M, E) \models V = L,\\ &(d)\ \pi : \text{“}(M, E) \cong L_\gamma\text{”} \text{ is the collapse map, and}\\ &(e)\ \pi(s) = S.\end{aligned}$$

So $S \in L$ and $G_{S,\mathscr{S}} \cap X \in L$, since $\mathscr{S}, X \in L$. Hence, by induction, $L \models G \cap X \in \mathrm{In}^\kappa_\gamma$. □

DEFINITION 26. Let M be a transitive standard model such that $M \models \mathrm{ZFC} + \kappa$ is a cardinal. $\mathscr{U}$ is an *M-ultrafilter* iff

(1) $(M, \mathscr{U}) \models \mathscr{U}$ is a κ-complete ultrafilter

(2) $(M, \mathscr{U})$ is *weakly amenable* in the sense that $\forall f \in {}^\kappa M \cap M\,(\{\zeta < \kappa \mid f(\zeta) \in \mathscr{U}\} \in M$.

$\mathscr{U}$ is a *normal M-ultrafilter* iff $\mathscr{U}$ is an M-ultrafilter such that $(M, \mathscr{U}) \models \mathscr{U}$ normal.

THEOREM 27. (Kleinberg) *Suppose M is a countable transitive standard model of* ZFC. *Then the following are equivalent*

(1) *There is a normal M-ultrafilter over κ*

(2) *$M \models \kappa$ is completely ineffable.*

Proof. (1) $\rightarrow$ (2): Suppose $\mathscr{U}$ is a normal M-ultrafilter on κ.

CLAIM. *$\mathscr{U} \subseteq (\mathrm{In}^\kappa_\beta)^M$ for all $\beta \in \Omega \cap M$.*

Proof. By induction on β. CASE 0: $\mathscr{U} \subseteq (\mathrm{In}^\kappa_0)^M$ since normality implies that M extends the club filter. CASE λ: Immediate. CASE $\gamma + 1$: Fix $X \in \mathscr{U}$. We have to

show that $X \in (\mathrm{In}^{\kappa}_{\gamma+1})^M$—that is, for each $\mathscr{S} \in M$ we must find an $S \in M$ such that

$$G_{S,\mathscr{S}} \cap X \in (\mathrm{In}^{\kappa}_{\gamma})^M.$$

So fix $\mathscr{S} \in M$. To find $\mathscr{S}$ consider the ultrapower ${}^{\kappa}M/\mathscr{U}$. Since $\mathscr{U}$ is an external measure we have no guarantee that ${}^{\kappa}M/\mathscr{U}$ is wellfounded but we do know that κ is a subset of the well-founded part of ${}^{\kappa}M/\mathscr{U}$ and that $[c_\alpha]_{\mathscr{U}} = \alpha$ for each $\alpha < \kappa$. Notice that for all $A \subseteq \kappa$ such that $A = [f]_{\mathscr{U}}$ we have

$$[c_\alpha]_{\mathscr{U}} \in [f]_{\mathscr{U}} \leftrightarrow \{\zeta < \kappa \mid c_\alpha(\zeta) \in f(\zeta)\} \in \mathscr{U}$$

and so

$$[f]_{\mathscr{U}} = \{\alpha < \kappa | \{\zeta < \kappa \mid \alpha \in f(\zeta)\} \in \mathscr{U}\}.$$

In particular,

$$[\mathscr{S}]_{\mathscr{U}} = \{\alpha < \kappa | \{\zeta < \kappa \mid \alpha \in \mathscr{S}(\zeta)\} \in \mathscr{U}\}.$$

Thus, $[\mathscr{S}]_{\mathscr{U}} \in M$ by weak amenability. Now take $S = [\mathscr{S}]_{\mathscr{U}}$. This works since, letting $A_\alpha = \{\zeta < \kappa \mid \alpha \in \mathscr{S}(\zeta)\}$ for each $\alpha \in S$, we have

$$\triangle_{\alpha \in S} A_\alpha \equiv \{\zeta < \kappa \mid \zeta \in \bigcap_{\alpha \in S \cap \zeta} A_\alpha\} = G_{S,\mathscr{S}}$$

which is in $\mathscr{U}$ by normality. □

Thus $(\mathrm{In}^{\kappa}_{\beta})^M \neq \varnothing$, for all $\beta \in \Omega \cap M$, and hence $M \models \kappa$ is completely ineffable.

(2) → (1): Suppose $M \models \kappa$ is completely ineffable. We seek a normal M-ultrafilter $\mathscr{U}$. We have a normal *M-filter*, viz., $\mathscr{F}_{\mathrm{ci}} \in M$. There is a standard way of extending such a filter to an ultrafilter: Let G be $\mathscr{I}^{+}_{\mathrm{ci}}$-generic over M where the ordering on $\mathscr{I}^{+}_{\mathrm{ci}}$ is given by: $q \leq p$ iff $q \setminus p \in \mathscr{I}_{\mathrm{ci}}$. G exists since M is countable. Standard arguments show that G is a normal M-ultrafilter, with the exception of *weak amenability*. Kleinberg's insight was to see how the combinatorial properties of the completely ineffable cardinal κ (in particular, the so-called *flipping* properties) could be leveraged to ensure weak amenability. To see that $\mathscr{F} \subseteq G$, fix $X \in \mathscr{F}$ and

notice that $\{Y \in \mathscr{I}^+_{\text{ci}} \mid Y \subseteq X\}$ is dense in $\mathscr{I}^+_{\text{ci}}$. To see that G is normal in the sense of M fix $X \in G$ and $f \in M$ such that $f : X \to \kappa$ is regressive and notice that $\{Y \subseteq X \mid f \text{ is constant on } Y\}$ is dense below X. Finally, we have to show that G is weakly amenable—that is, for all $f \in {}^{\kappa}M \cap M$ $(\{\zeta < \kappa \mid f(\zeta) \in G\} \in M)$. To this end we make the following definitions. Suppose $f, g : \kappa \to \mathscr{P}(\kappa)$. The sequence g is a *flip* of f iff for all $\alpha < \kappa$, $g(\alpha) \in \{f(\alpha), \kappa \setminus f(\alpha)\}$. The *diagonal intersection of* f is $D_g \equiv \{\alpha < \kappa \mid \forall \zeta < \alpha \ (\alpha \in f(\zeta))\}$. Now fix $f \in {}^{\kappa}M \cap M$. We have to show that

$$\{Y \in \mathscr{I}^+_{\text{ci}} \mid \exists g \sim f \ (Y \subseteq D_g)\}$$

is dense.

CLAIM. *For each* $f : \kappa \to \mathscr{P}(\kappa)$ *and* $X \in \mathscr{I}^+_{\text{ci}}$ *there is a* $g \sim f$ *and a* $Y \in \mathscr{I}^+_{\text{ci}}$ *such that* $Y \subseteq X \cap D_g$.

Proof. Define $\mathscr{S}$ such that

$$\mathscr{S}(\alpha) = \{\zeta < \alpha \mid \alpha \in f(\zeta)\}$$

and by the complete ineffability of κ choose S such that $G_{S,\mathscr{S}} \cap X \in \mathscr{I}^+_{\text{ci}}$. Let $g \in M$ be such that

$$g(\zeta) = \begin{cases} f(\zeta) & \text{if } \zeta \in S \\ \kappa \setminus f(\zeta) & \text{otherwise} \end{cases}$$

Then $Y \equiv G_{S,\mathscr{S}} \cap X \subseteq D_g \cap X$, as if $\alpha \in G_{S,\mathscr{S}}$ then we have $\alpha \in f(\zeta) = g(\zeta)$ for each $\zeta \in S \cap \alpha$ and $\alpha \in \kappa \setminus f(\zeta) = g(\alpha)$ for each $\zeta \in \alpha \setminus S$. □

This completes the proof of the theorem. □

COROLLARY 28. *Assume* $0^{\#}$ *exists. Then* $L \models \iota_\zeta$ *is completely ineffable, for each Silver indiscernible* ι_ζ.

Proof. Let j be an order preserving map from Silver indiscernibles to Silver indiscernibles with critical point ι_ζ. This map uniquely extends to an elementary embed-

ding $\tilde{\jmath} : L \to L$ with $\mathrm{crit}(\tilde{\jmath}) = \iota_\zeta$. Set

$$\mathscr{U} = \{X \subseteq \iota_\zeta \mid X \in L \wedge \iota_\zeta \in \tilde{\jmath}(X)\}.$$

It is straightforward to see that $\mathscr{U}$ is a normal L-ultrafilter. Thus, by Kleinberg's theorem, $L \models \iota_\zeta$ is completely ineffable. □

COROLLARY 29. *Suppose that $\kappa(\omega)$ exists. Then there is a cardinal less than $\kappa(\omega)$ which is completely ineffable and hence n-reflective for all n.*

Proof. Let $\kappa = \kappa(\omega)$. Following the proof of Silver-Reinhardt that there is a totally indescribable cardinal less than κ we consider the structure $\mathscr{M} = (V_\kappa, \in, <)$ where $<$ well-orders V_κ. Let $I = \{\iota_n\} \in \kappa$ be the indiscernibles of $\mathscr{M}$ given by Silver's theorem. Take the Skolem hull of these indiscernibles and collapse, letting

$$\begin{aligned} \mathscr{H} &= \mathrm{Hull}^{\mathscr{M}}(I), \\ \pi : \mathscr{H} \cong \bar{\mathscr{H}} &= \mathrm{collapse}(\mathscr{H}), \text{and} \\ \bar{I} &= \pi(I). \end{aligned}$$

Notice that $\bar{I}$ is a set of indiscernibles for $\bar{\mathscr{H}}$ and that by including the well-ordering $<$ in $\mathscr{M}$ we ensure that these indiscernibles, which we will enumerate as $\{\bar{\iota}_n\}$, obey the key properties (with respect to $\bar{\mathscr{H}}$) which are obeyed by the Silver indiscernibles (with respect to L).

Now let $j : \bar{I} \to \bar{I}$ be order preserving and such that $\mathrm{crit}(j) = \bar{\iota}_n$. This map uniquely extends to an elementary embedding $\tilde{\jmath} : \bar{\mathscr{H}} \to \bar{\mathscr{H}}$ with $\mathrm{crit}(\tilde{\jmath}) = \bar{\iota}_n$. Set

$$\mathscr{U} = \{X \subseteq \bar{\iota}_n \mid X \in \bar{\mathscr{H}} \wedge \bar{\iota}_n \in \tilde{\jmath}(X)\}.$$

It is straightforward to see that $\mathscr{U}$ is a normal $\bar{\mathscr{H}}$-ultrafilter. Thus, by Kleinberg's theorem, $\bar{\mathscr{H}} \models \iota_n$ is completely ineffable. So we have $\mathscr{H} \models \iota_n$ is completely ineffable, $\mathscr{M} \models \iota_n$ is completely ineffable, and finally, since κ is inaccessible, $V \models \iota_n$ is completely ineffable. □

Chapter 4

Extension Principles

In this chapter I will introduce a new series of principles—*extension principles*—and show that these principles have much greater scope than reflection principles. The main result is that unlike reflection principles, extension principles can overcome the Minimal Hurdle (and much more), thereby effecting a significant reduction in incompleteness. I will begin in §4.1 by discussing an approach to the justification of large cardinal axioms due to Reinhardt. I will then show that this approach has two problems—the problem of *tracking* and the problem of *extendibility to inconsistency*. Extension principles are designed to overcome both of these problems. I will introduce one of the weaker extension principles, EP, in §4.2 and show that it overcomes the problem of tracking. In §4.3 I will show that EP overcomes the problem of extendibility to inconsistency and that it overcomes the Minimal Hurdle. I will also isolate the exact strength of EP. §4.4 is devoted to the much stronger extension principle GEP. The main result is that GEP implies PD and freezes the theory of second order arithmetic.

4.1 Reinhardt

4.1.1 Introduction. Reinhardt's approach to the justification of large cardinal axioms is to take *the totality of sets* V as a additional primitive and then investigate and characterise the properties of this primitive. In terms of our earlier distinction

between axioms of nature and axioms of extent Reinhardt's approach is therefore to investigate axioms of nature, but axioms which pertain the nature of the *totality* of sets and not to the nature of *individual* sets.

I will focus on the following three sources: Reinhardt's thesis, *Topics in the Metamathematics of Set Theory* (1967); an unpublished note, "Some strong axioms of infinity" (1968); and the paper "Remarks on reflection principles, large cardinal, and elementary embeddings" (1974).

The bulk of Reinhardt's thesis is concerned with the system of set theory introduced by Ackermann, the main result being that Ackermann's set theory A* is equivalent to ZF. (Levy had already proved the forward implication). My primary interest, however, is in the final chapter of the thesis where Reinhardt uses the motivation underlying A* to justify a system ZE which, in modern terminology, yields the existence of a 1-extendible cardinal. There is a single sentence in the thesis (almost a parenthetical remark) which indicates how to extend this system to a much stronger system. It is this sentence which has made the thesis famous. For, assuming the axiom of choice, the proposed extension is inconsistent.

4.1.2 Ackermann Set Theory. Ackermann's set theory with foundation, A*, consists of the following axioms in the language of set theory with a constant V: Extensionality, Separation restricted to V, Foundation, Closure of V under $\in$ and $\subseteq$, and Ackermann's principle:

> If $\{x \mid \varphi(x, a, b)\} \subseteq V$ (where $a, b \in V$ and 'V' *does not* occur in φ) then $\{x \mid \varphi(x, a, b)\} = c$ for some $c \in V$.

The theory A is the theory A* without Foundation. The motivation for Ackermann's principle is the idea that any "sharply delimited" collection is a set. The antecedent of the principle is supposed to give a sufficient condition for being sharply delimited, i.e. if a collection C of sets is definable with parameters without reference to V, then C is sharply delimited. I will spell this out in some detail in the next section but first it will be instructive to state the principal results concerning A and A*

THEOREM 30. (Levy) *If* $\mathrm{A}^* \vdash \varphi^V$ *then* $\mathrm{ZF} \vdash \varphi$.

THEOREM 31. (Reinhardt) *If* $\mathrm{ZF} \vdash \varphi$ *then* $\mathrm{A}^* \vdash \varphi^V$.

THEOREM 32. (Reinhardt) *Assume* A. *Suppose* ψ *is a formula and* Φ *is a finite set of formulas, each with one free variable. If* $\psi(V)$ *then there is a* V' *such that*

(1) $V \in V'$,

(2) $\psi(V')$, *and*

(3) $\Phi^V(x) \leftrightarrow \Phi^{V'}(x)$ *for all* $x \in V$.

THEOREM 33. (Reinhardt) *Suppose* $V_{\alpha+1} \prec V_{\beta+1}$ *where* $\alpha < \beta$. *Then*

(1) $(V_{\beta+1}, \in, V_\alpha) \models \mathrm{A}$ *and*

(2) *If* $V = L$, *then* α *is* Π^1_ω*-indescribable.*

4.1.3 The Principle of Sharp Delimitation. With these technical results behind us let us now turn to the motivation for Ackermann's principle.

Ackermann thought that in order for a collection to be a set it must be the case that "what belongs to the collection and what does not belong to it must be sufficiently sharply delimited" Ackermann (1956, p. 282). Since the collection of all sets is not a set, it follows that on Ackermann's view the notion of set is not "sharply delimited". But what is meant by 'sharply delimited'? The antecedent of Ackermann's principle provides a precise sufficient condition, namely, if $\{x \mid \varphi(x, a, b)\} \subseteq V$ (where $a, b \in V$ and 'V' *does not* occur in φ) then $\{x \mid \varphi(x, a, b)\}$ is sharply delimited and hence by the principle is a set. One way to think of the motivation for this is along the lines of the potentialist conception introduced earlier. On the potentialist view we have been considering there are many legitimate candidates for V. Let us enumerate the spectrum of the concept of set as $\Lambda = \{\kappa_\alpha \mid \alpha \in \Omega\}$. Thus, for each $\kappa_\alpha \in \Lambda$, the structure V_{κ_α} meets the concept of set and so the structure V_{κ_α} satisfies all of the closure properties inherent in the concept of set. As noted earlier no one candidate is better than another. The idea is that the legitimate candidates are order

indiscernibles. The constant 'V' is to be taken as a variable that ranges over the various candidates and there is no clear choice as to which candidate is the best one—they are all on a par, they are indiscernible. The antecedent of Ackermann's principle says that if a collection C is definable with parameters in V but without reference to V—that is, if the interpretation of the (intensional) collection C is independent of the actual reference of 'V'—then the collection C is a set.

We have still not given a precise definition of 'sharp delimitation', although at the close of the last paragraph we gave a rough formulation, namely, a collection is sharply delimited if and only if its interpretation is independent of the interpretation of 'V'. Let us make this more precise, following Reinhardt pp. 72–74. Let $\varphi(x)$ be a formula with one free variable. It defines a collection over the background universe which contains (the various candidates for V). *Relative* to a given candidate V this collection corresponds to $\varphi^V = \{x \mid \varphi(x) \wedge x \in V\}$. A collection is said to be *sharply delimited* if and only if φ^V is independent of the interpretation of 'V', that is, if and only if,

$$\forall \alpha, \beta \ (\varphi^{V_{\kappa\alpha}} = \varphi^{V_{\kappa\beta}}).$$

So the principle which says that 'if a collection C of sets is sharply delimited then C is a set' can be written as

$$\forall \alpha, \beta \ (\varphi^{V_{\kappa\alpha}} = \varphi^{V_{\kappa\beta}}) \rightarrow \exists c \in V_{\kappa_0} \ (c = \varphi^{V_{\kappa_0}}).$$

Let us call this principle *the principle of sharp delimitation.* Notice that this principle implies Ackermann's principle, since the antecedent of Ackermann's principle is a sufficient condition for sharp delimitation. But sharp delimitation is broader than Ackermann's sufficient condition, and so it is natural to consider the system of set theory based on the stronger principle of sharp delimitation.

DEFINITION 34. The system of set theory A^+ consists of the following axioms in the language of set theory with a constants V and V':

(1) The axioms of A^* except Ackermann's principle

(2) $V \subseteq V'$

(3) $\varphi^V = \varphi^{V'} \rightarrow \exists c \in V\ (c = \varphi^V)$

It is easy to see that one can obtain a model of A^+ from a Π^1_ω-indescribable cardinal.

4.1.4 Comparing the Candidates. Let V_{κ_0}, V_{κ_1}, V_{κ_2}, V_{κ_4} be the first four candidates for V. These are regarded as the possible worlds of set theory, the legitimate candidates. Working in V_{κ_4} we would like to examine the structure $(V_{\kappa_2}, \in, V_{\kappa_1}, V_{\kappa_0})$ and determine what must be true of it if V_{κ_1} and V_{κ_0} are to serve as legitimate candidates for V. Reinhardt's basic method for obtaining reflection principles "is to exploit the principle which says that mathematical truths should be necessary truths" and "[a]ccording to this principle, if the notion of possibility we have introduced is a good one, something true in one interpretation of V should be necessarily true, that is, true in all possible alternative interpretations of V" (p. 76).

This principle is made more precise by specifying the language. First, consider the standard first-order language of set theory, that is, $\mathscr{L}_{1,0}$. In this language the principle that mathematical truths are necessary truths is expressed by the schema

$$\varphi^{V_{\kappa_0}} \leftrightarrow \varphi^{V_{\kappa_1}}$$

In other words, the principle asserts that $V_{\kappa_0} \equiv V_{\kappa_1}$. Second, consider the language $\mathscr{L}_{1,1}$, that is the first-order language of set theory *with parameters*. When this language is interpreted over V_{κ_α} the parameters are the elements of V_{κ_α}. In this language the principle that mathematical truths are necessary truths is expressed by the schema

$$\forall x \in V_{\kappa_0}\ (\varphi^{V_{\kappa_0}}(x) \leftrightarrow \varphi^{V_{\kappa_1}}(x)).$$

In other words, the principle asserts that $V_{\kappa_0} \prec V_{\kappa_1}$. One can justify the above schema without reference to the notion of possibility by noting that if V_{κ_0} and V_{κ_1} are really legitimate candidates for V then they had better satisfy the same sentences in as rich a language as possible (without, of course, making actual reference to V_{κ_0} or

V_{κ_1}). This avoids the difficulties surrounding how one is to interpret possibility in the mathematical context where, it would seem, not only are mathematical *truths* necessary but also mathematical *objects* exist necessarily.

Notice that these principles are similar to reflection principles. The main difference so far is that the all of the statements true of V_{κ_1} are being reflected to a *fixed* level, namely V_{κ_0}. A second difference will become clear when we enrich the language to include higher-order parameters.

The next step is to consider the languages $\mathscr{L}_{2,1}$, $\mathscr{L}_{3,1}$, ..., $\mathscr{L}_{\beta,1}$, ... Doing so does not lead to a significant improvement in strength. The significant jump occurs when we try to use the language $\mathscr{L}_{2,2}$, that is, the language of second-order set theory with second-order parameters. This is what Reinhardt does at the close of his thesis.

A problem arises when dealing with sentences in $\mathscr{L}_{2,2}$. Recall that in the case of $\mathscr{L}_{1,1}$ the parameters were elements of V_{κ_0} and a statement true of a parameter a over V_{κ_0} was true of that *very same parameter* a over V_{κ_1}. But if A is a second-order parameter over V_{κ_0}—that is, a subclass of V_{κ_0}—then this will fail; for example, if A is unbounded in V_{κ_0} then the sentence 'A is unbounded' will be true when interpreted over V_{κ_0} but not when interpreted over V_{κ_1}. Thus to each unbounded subclass A of V_{κ_0} we need to *associate* an unbounded subclass $j(A)$ of V_{κ_1}. It is helpful to think of the first-order parameters a having names c_a and the second-order parameters A having names C_A. Then the issue is how the name C_A is to be interpreted over V_{κ_1}. Reinhardt outright assumes that there is an interpretation function (equivalently, a map j) which is such that

$$\forall x \in V_{\kappa_0}\, \forall X \subseteq V_{\kappa_0}\, \big(\varphi^{V_{\kappa_0}}(x, X) \leftrightarrow \varphi^{V_{\kappa_1}}(x, j(X))\big).$$

This principle asserts that $V_{\kappa_0+1} \prec V_{\kappa_1+1}$, where κ_0 is inaccessible, that is, the principle asserts that κ_0 is 1-extendible.

4.1.5 Criticism. There are two problems.

The first problem is that we are not told what $j(A)$ is and we are given no reason for thinking that there is a way of associating a $j(A)$ to each A in such a way that the above schema comes out true. We are given no guide as to how to track parameters from one structure to another. Let us call this *the problem of tracking.*

The second problem is that the above schema leads to inconsistency when generalised. "[I]n order to extend [the above schema] to allow parameters of arbitrary (in the sense of V_{κ_2} order over V_{κ_0}) we simply remove the restriction '$X \subseteq V_{\kappa_0}$'" (p. 79; I have changed Reinhardt's 'V_0' to our 'V_{κ_0}' and so on.) The point is that there is nothing stopping us from doing this, that is, the justification for a 1-extendible is also justification for the result of making the above generalisation. The trouble is that when one does this the result is an elementary embedding $j : V \to V$ which, as Kunen showed, is inconsistent (with AC).[6] By the time of his (1974) Reinhardt was aware of Kunen's result and he refrains from stating the inconsistent axiom. But the point still remains that the argument he gives for a 1-extendible is also an argument for the inconsistent axiom. He has not drawn a principled distinction between the two. Let us call this *the problem of extendibility to inconsistency.*

4.2 The Principle EP

4.2.1 Tracking Definable Subsets. There is a natural way to overcome the problem of tracking. Consider two legitimate candidates V' and V'' for V where the former is properly contained in the latter. The problem of tracking is that given an *arbitrary* subclass A of V' there is no way of isolating the "corresponding" subclass $j(A)$ of V''. Notice, however, that if A is *definable* over V' via the formula φ and parameters $a_1 \cdots a_n$ then there is a way of isolating the corresponding subset $j(A)$ of V'', namely by implementing the definition of A over the structure V''—that is, by setting $j(A) = \{x \in V'' \mid V'' \models \varphi(x, a_1 \ldots a_n)\}$.

On the potentialist view that we are pursuing there are many legitimate candidates which meet the concept of set. What is V according to the potentialist? It is some V_{κ_α}

which meets the concept of set. Thus, for each $\kappa_\alpha \in \Lambda$, the structure V_{κ_α} meets the concept of set and so the structure V_{κ_α} satisfies all of the closure properties inherent in the concept of set. Let us add two constants 'V'' and 'V''' to the language of set theory, each of which is intended to denote an element of the spectrum, say $V' = V_{\kappa_\alpha}$ and $V'' = V_{\kappa_\beta}$ where $\kappa_\alpha < \kappa_\beta$.

Since each of V' and V'' meets the concept of set we know that each is undefinable, something we made precise in terms of reflection principles. Thus any (higher-order) statement true of, say, V' reflects to some smaller V_α. But we know more than that V' and V'' are undefinable—we know that V' and V'' are *indistinguishable* since each is a permissible candidate for V, that is, each meets the concept of set. Let me spell out in detail what it means to say that V' and V'' are indistinguishable.

4.2.2 The Principle. We are given two structures, V' and V'', each of which meets the concept of set and such that $V' \subseteq V''$. The principle EP can be reached in eight steps: (1) V' and V'' are both models of ZFC. (2) V' and V'' satisfy exactly the same sentences in the language of set theory. For if this were *not* true then there would be a sentence φ true in one structure but not in the other and this sentence would enable us to tell the two structures apart, thereby indicating that the two structures are not *equally good* candidates for V. (3) For exactly the same reason V' and V'' satisfy exactly the same sentences in the language of set theory even when we allow constants for sets common to both structures. Thus $V' \prec V''$. (4) Let X be a subclass of V' which is defined in V' by a formula ψ with parameters $a_1 \cdots a_n$. Let $j(X)$ be the subclass of V'' which is defined in V'' by the same formula ψ and the same parameters $a_1 \cdots a_n$. For example, ψ might define "the ordinals". In this case, X consists of the ordinals from the point of view of V' (that is, the elements of κ_α) and $j(X)$ consists of the ordinals from the point of view of V'' (that is, the elements of κ_β). Reasoning as above, we have V' *along with* X and V'' *along with* $j(X)$ satisfy exactly the same sentences, that is, $(V', X) \prec (V'', j(X))$. (5) Let $\mathrm{Def}(W)$ be the set

of all subsets of M definable in the structure W with parameters in W. Let

$$L_0(W) = W$$
$$L_{\alpha+1}(W) = \mathrm{Def}(L_\alpha(W))$$
$$L_\lambda(W) = \bigcup_{\alpha<\lambda} L_\alpha(W)$$

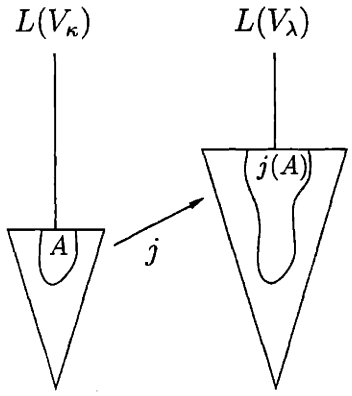
where α is a successor ordinal and λ is a limit ordinal. Finally, let $L(W) = \bigcup_{\alpha<\Omega} L_\alpha(W)$. Then the map $j : L_1(V') \to L_1(V'')$ must be an *elementary embedding* for otherwise we should be able to distinguish V' and V''. (6) We can now consider definable classes X over $L_1(V')$ and their analogues $j(X)$ over $L_1(V'')$. Reasoning as above, we have that $j : L_2(V') \to L_2(V'')$ is an *elementary embedding.* (7) Generalising we have that $j : L(V') \to L(V'')$ is an *elementary embedding.* (8) Finally, putting all of this together we have: For each $\kappa, \lambda \in \Lambda$ such that $\kappa < \lambda$ there is an elementary embedding $j_{\alpha\beta} : L(V_\kappa) \to L(V_\lambda)$. As our official principle let us take the following:

EP. *For each α there is an elementary embedding*

$$j : L(V_\kappa) \to L(V_\lambda)$$

such that $\alpha < \kappa < \lambda$.

Here's the picture:



According to the potentialist view we are pursuing, the permissible candidates V_κ for V are equivalent in a sense partially fleshed out by EP.

4.3 The Strength of EP

4.3.1 Freezing. The following result is a straightforward consequence of the well-known construction for $\underset{\sim}{\prod}^1_1$-sets due to Martin and Solovay.

THEOREM 35. EP *freezes* $\underset{\sim}{\sum}^1_3$.

Proof. First we note that EP implies that $X^\#$ exists for all sets X: Fix α such that $X \in V_\alpha$. By EP there is a non-trivial elementary embedding

$$j_{\kappa\lambda} : L(V_\kappa) \to L(V_\lambda)$$

where $\alpha < \kappa < \lambda$. The restriction of this map to $L(X)$,

$$j_{\kappa\lambda}|X : L(X) \to L(X)$$

is an elementary embedding. Thus $X^\#$ exists.

Now let $G \subseteq \mathbb{P}$ be V-generic. Choose $\kappa > \mathrm{Card}(\mathbb{P})$. EP implies that $(V_\kappa)^\#$ exists. Let T be the Martin-Solovay tree for a $\underset{\sim}{\prod}^1_2$-set defined by φ, constructed relative to $(V_\kappa)^\#$. (See Kanamori (1997, Ch. 15) for details, using the sharp in place of the measurable cardinal.) Since $\mathbb{P}$ is a small forcing relative to κ the embeddings in the Martin-Solovay tree lift to embeddings for the corresponding tree in $V[G]$ (by the proof of the central theorem in Levy & Solovay (1967). Thus $T^G = T$. Now suppose $V \models \exists x\, \varphi(x)$. Then $x \in p[T] \subseteq p[T]^{V[G]}$ and so $V[G] \models \exists x\, \varphi(x)$. Conversely, suppose that $V \models \neg\exists x\, \varphi(x)$. Then $T = T^{V[G]}$ is well-founded and so $V[G] \models \neg\exists x\, \varphi(x)$. □

Thus in the presence of EP on cannot alter the $\underset{\sim}{\sum}^1_3$-theory via forcing: EP ensures that as far as $\underset{\sim}{\sum}^1_3$ sentences are concerned consistency via forcing implies truth. We have finally reached a justified principle which overcomes the Minimal Hurdle and effects a significant reduction in incompleteness. In particular, EP implies SPHERE and lifts all of the results of classical descriptive set theory to the next projective level.

4.3.2 Sharps. One would like further assurance that EP is consistent. It is easy to see that 'ZFC + There is a proper class of superstrong cardinals' implies EP. Perhaps this is not very reassuring. Fortunately, one can prove a much sharper bound. In the next section I will isolate the optimal bound: EP is equivalent to the statement 'for all X, $X^{\#}$ exists'. This will have an interesting philosophical consequence which I will discuss in §4.3.4. The purpose of the present section is to set up the requisite machinery.

There are many equivalent formulations of the statement that $X^{\#}$ exists. For our purposes the inner model theoretic formulation will be most useful. In what follows I will outline the theory without giving proofs. The interested reader is referred to Steel (1999) and Schimmerling (2001).

Assume $X^{\#}$ exists, i.e. there is a non-trivial elementary embedding $j : L(X) \to L(X)$. Our aim is to construct a countable model $\mathscr{M}$ which contains all of the information inherent in j. We proceed in two steps. The first step is to find a *set* which contains the information of the class j. The second step will be to find a set of cardinality $\mathrm{Card}(X)$ which encodes the same information.

The first step. Let κ be the critical point of j. It is easy to see that

$$U = \{A \subseteq \kappa \mid A \in L(X) \wedge \kappa \in j(A)\}$$

is a normal ultrafilter over $L(X)$. Let us say that a structure $(M, \in, A)$ is *amenable* if and only $A \cap x \in M$ for all $x \in M$. For example, the structure $M = (L(X), \in, U)$ is not amenable since $U \notin M$. Moreover, the longest initial segment of $L(X)$ which has a chance of being amenable is $L_{(\kappa^+)^M}(X)$. Kunen showed that $(L_{(\kappa^+)^M}(X), \in, U)$ is indeed amenable. Let us introduce some more terminology.

DEFINITION 36. An *$X^{\#}$-premouse* is a structure $\mathscr{M} = (L_\lambda(X), \in, U)$ such that

(1) $\lambda = (\kappa^+)^{L(X)}$,

(2) $L_\lambda(X) \models \mathrm{ZF} - \mathrm{Powerset} + $'$\kappa$ is the largest cardinal',

(3) $\mathscr{M} \models$'U is a normal ultrafilter over κ', and

(4) $\mathscr{M}$ is amenable.

Using this terminology we can summarise the situation as follows: Kunen showed that if $X^{\#}$ exists then there is an $X^{\#}$-premouse.

Let us continue with our assumption that $X^{\#}$ exists and see what further information we can extract. We work in a language that contains a constant $\dot{a}$ for each $a \in X$. The goal, remember, it to extract enough information to enable us to recover $X^{\#}$. Given an $X^{\#}$-premouse $\mathscr{M} = (L_\lambda(X), \in, U)$ we can take the ultrapower of $\mathscr{M}$ by U. Assuming the ultrapower is well-founded we can identify it with its transitive collapse. It is straightforward to show that Loś' theorem holds in this setting and that the canonical map $i_U(x) \mapsto [c_x]_U$ (where $c_x : \alpha \mapsto x$ is the constant function with domain κ) is cofinal in the ultrapower. Furthermore, setting $U' = \bigcup_{\alpha\in\lambda} i_U(U \cap L_\alpha(X))$ we have that the ultrapower $\mathscr{M}' = (L_{\lambda'}(X), \in, U')$ is an $X^{\#}$-premouse. Thus we can take the ultrapower of $\mathscr{M}'$ by U' and continue in this fashion, so long as the ultrapower at each stage is well-founded. At limit stages of the process we take the direct limit. Let us write $\mathscr{M}_\alpha$ for the α^{th} model of this iteration of $\mathscr{M}$ and call it the α^{th} *iterate of* $\mathscr{M}$. $\mathscr{M}$ is *α-iterable* if for all $\beta < \alpha$ if $\mathrm{Ult}(\mathscr{M}_{\bar{\beta}}, U_{\bar{\beta}})$ is well-founded when $\beta = \bar{\beta} + 1$ and the direct limit of the models $(\mathscr{M}_{\bar{\beta}})_{\bar{\beta}<\beta}$ is well-founded when β is a limit ordinal. We say that $\mathscr{M}$ is *iterable* iff $\mathscr{M}$ is α-iterable for all $\alpha < \Omega$ and we say that $\mathscr{M}$ is an *$X^{\#}$-mouse* iff it is *iterable.*

Now one can show that if $X^{\#}$ exists then there is an $X^{\#}$-mouse. (The point is that $\mathscr{M}^* = (L(X), \in, U)$ is iterable and the models of the $\mathscr{M}$-iteration embed into the corresponding models of the $\mathscr{M}^*$-iteration.) Thus we have an Ω-length iteration of $\mathscr{M}$. We can now recover $X^{\#}$. The limit model of this iteration is $L(X)$ and the critical points of the embeddings are $L(X)$-indiscernibles. But any order preserving map from the $L(X)$-indiscernibles into themselves induces an elementary embedding from $L(X)$ into itself, which gives us $X^{\#}$. The completes the first step of the analysis: we have a set-sized structure which encodes $X^{\#}$.

Second step: Suppose $\mathscr{M}$ is an $X^{\#}$-premouse. Let $\mathscr{L}_U$ be the language of set theory $\mathscr{L}$ with an additional constant symbol $\dot{U}$. For $Y \subseteq M$ let $\mathrm{Hull}_1^{\mathscr{M}}(Y)$ consist of all elements $a \in M$ such that a is the unique element x such that $\mathscr{M} \models \varphi \wedge \psi[x, \bar{b}]$

where φ is a Σ_1 formula of $\mathscr{L}_U$, ψ is a formula of $\mathscr{L}$, and $\bar{b}$ is a sequence of parameters from Y. We have that $\mathrm{Hull}_1^{\mathscr{M}}(Y)$ is a Σ_1-substructure of $\mathscr{M}$ with respect to $\mathscr{L}_U$ and a fully elementary substructure of $\mathscr{M}$ with respect to $\mathscr{L}$. Let us call an $X^{\#}$ premouse $\mathscr{M}$ *minimal* iff $\mathscr{M} = \mathrm{Hull}_1^{\mathscr{M}}(\varnothing)$. The key fact is the following: if there is an $X^{\#}$-mouse then there is a *unique* minimal $X^{\#}$-mouse. We let $\mathscr{M}_0^{\#}(X)$ denote the minimal $X^{\#}$-mouse. Note that $\mathscr{M}_0^{\#}(X)$ is countable and that its Ω-length iteration gives rise to indiscernibles which yield an elementary embedding from $L(X)$ into itself. Thus we have outlined the proof of the following:

FACT 37. *The following are equivalent*:

(1) *There is a non-trivial elementary embedding* $j : L(X) \to L(X)$.

(2) $\mathscr{M}_0^{\#}(X)$ *exists.*

Thus the class size embedding which constitutes $X^{\#}$ can be represented as the countable iterable inner model $\mathscr{M}_0^{\#}(X)$.

4.3.3 Strength. We are now in a position to pinpoint the exact strength of EP.[7]

THEOREM 38. *Assume that there is a proper class of inaccessibles. Then the following are equivalent*:

(1) EP.

(2) *For all* X, $X^{\#}$ *exists.*

Proof. (1) $\to$ (2): As above. (Choose α such that $X \in V_\alpha$. Then by EP there is an elementary embedding $j_{\kappa\lambda} : V_\kappa \to V_\lambda$ where $\alpha < \kappa < \lambda$. The restriction of this embedding to $L(X)$ witnesses that $X^{\#}$ exists.)

(2) $\to$ (1): Assume that $X^{\#}$ exists for all X. Let I be the proper class of inaccessibles. Fix α. We seek κ, λ and $j_{\kappa\lambda}$ such that $\alpha < \kappa < \lambda$ and $j_{\kappa\lambda} : L(V_\kappa) \to L(V_\lambda)$ is an elementary embedding. Start by letting λ be an inaccessible greater than α. We have that $(V_\lambda)^{\#}$ exists and hence by the results of the previous section that

$\mathscr{M}_0 = \mathscr{M}_0^{\#}(V_\lambda) = (L_\gamma(V_\lambda), \in, U)$ exists. Let

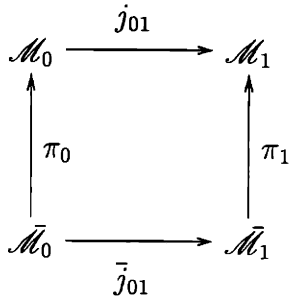
$$\begin{aligned} H_0 &= \mathrm{Hull}_1^{\mathscr{M}_0}(V_\alpha) & \kappa_0 &= \sup(H_0 \cap V_\lambda) \\ H_{n+1} &= \mathrm{Hull}_1^{\mathscr{M}_0}(H_n \cup V_{\kappa_n}) & \kappa_{n+1} &= \sup(H_{n+1} \cap V_\lambda) \\ H_\omega &= \bigcup_{n<\omega} H_n & \kappa &= \bigcup_{n<\omega} \kappa_n \end{aligned}$$

Now let $\bar{\mathscr{M}}_0 = \mathrm{collapse}(H_\omega) = (L_{\bar{\gamma}}(V_\kappa), \in, \bar{U})$ and let $\pi_0 : \bar{\mathscr{M}}_0 \to \mathscr{M}_0$ be the inverse of the collapse map. Note that $\pi(\kappa) = \lambda$ and that $\bar{\mathscr{M}}_0$ is a $(V_\kappa)^{\#}$-premouse.

Let $(\mathscr{M}_\alpha)_{\alpha<\Omega}$ be the iteration of $\mathscr{M}$ and let $j_{\alpha\beta} : \mathscr{M}_\alpha \to \mathscr{M}_\beta$ be the iteration maps. Let $\mathscr{M}_\infty$ be the limit model and note that $\mathscr{M}_\infty = L(V_\lambda)$. Let $\bar{\mathscr{M}}_1 = \mathrm{Ult}(\bar{\mathscr{M}}_0, \bar{U})$ and let $\bar{j}_{01}$ be the ultrapower map. It is easy to check that if we let

$$\begin{aligned} \pi_1 : \bar{\mathscr{M}}_1 &\to \mathscr{M}_1 \\ [f]_{\bar{U}}^{\bar{\mathscr{M}}_0} &\mapsto [j_{01}(f)]_U^{\mathscr{M}} \end{aligned}$$

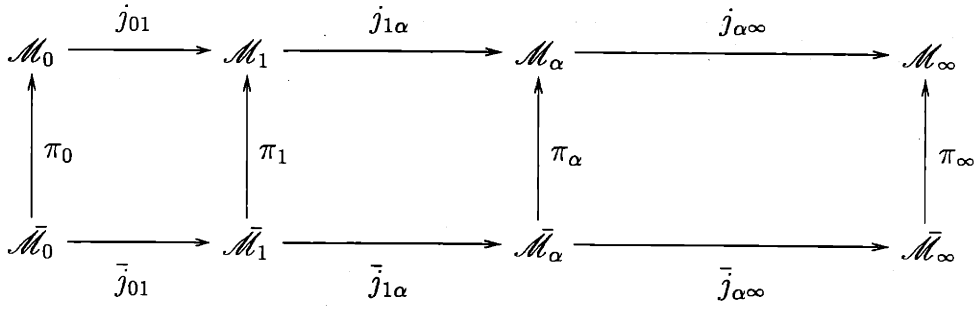
where $f : \mathrm{crit}(\bar{U}) \to \bar{M}$ is a function in $\bar{M}_0$, then the following diagram



commutes and all of the embeddings are elementary with respect to $\mathscr{L}$. Continuing in this fashion, supposing that we have obtained π_α, we let

$$\begin{aligned} \pi_{\alpha+1} : \bar{\mathscr{M}}_{\alpha+1} &\to \mathscr{M}_{\alpha+1} \\ [f]_{\bar{U}}^{\bar{\mathscr{M}}_\alpha} &\mapsto [j_{\alpha\alpha+1}(f)]_{U_\alpha}^{\mathscr{M}} \end{aligned}$$

and take direct limits at limit stages. We thus obtain the following directed system:



The direct limit map $\pi_\infty : L(V_\kappa) \to L(V_\lambda)$ is our desired elementary embedding. □

4.3.4 Inevitability. Our justification of the principle EP has its roots in the potentialist view of the universe of sets. The potentialist takes seriously the view that any given kind or plurality of objects can be formed into a set. This has the consequence that "the sets" do not form a given kind. Nevertheless it seems that we can make meaningful statements about the concept of set or the totality of sets. There are a number of ways of explaining this. One way is through the device of typical ambiguity. Another is to regard V as an order indiscernible, that is, to regard talk of V as talk of one of (proper class-) many indistinguishable legitimate candidates for V. On this view when we talk of V we are talking about a set. True we can go further and take the powerset of this set but in doing so we do not obtain an *essentially* richer conception. Of course, we do obtain a *trivially* richer conception, trivial in the sense that $V_{\kappa+1}$ is richer than V_κ. But we will never (provided we avoid reference to V) reach a level which has closure conditions that surpass those of V. This is what it means to say that we never obtain an essentially richer conception. I present this view as an option. It is beyond the scope of this thesis to provide a full defense of the view.

There is another justification of EP. In addition to being justified by the aforementioned conception of set this principle is in a certain sense inevitable. The purpose of this subsection is to explain the sense in which this is true.

Recall the notion of a *generically invariant set of axioms* that was introduced in

§3.4.1: A is *generically invariant* if and only if ZF + A is preserved under set-size forcing extensions. For example, 'There is an inaccessible cardinal' is not generically invariant but 'There is a proper class of inaccessible cardinals' is generically invariant. All of the new axioms we have considered thus far (that is, reflection principles and extension principles) are generically invariant. (More generally, we have been investigating axioms which assert that there are large rank initial segments of V with certain largeness properties P. Suppose T is a theory which yields a rank initial segment of V with the largeness property P. The theory T has generically invariant extension T′ asserting that there is a proper class of rank initial segments with the largeness property V. Now our concern is with whether extensions T of ZFC make a significant reduction in incompleteness. Since the stronger the theory the greater the reduction in incompleteness we can, without loss of generality, focus on theories which are generically invariant.)

Now recall Corollary 20 from Chapter 3 which states that if A is generically invariant and ZF+A freezes $\underset{\sim}{\Sigma}^1_3$ then $X^\#$ exists for all X. Given this and the theorem above we have:

COROLLARY 39. *If* A *is generically invariant and* ZF + A *freezes* $\underset{\sim}{\Sigma}^1_3$ *then* EP *holds.*

In this sense the principle EP is inevitable: it is implied by *any* generically invariant theory which overcomes the Minimal Hurdle.

4.4 The Principle GEP

4.4.1 The Generalised Extension Principle. In overcoming the problem of tracking we restricted our attention from V to the inner model $L(X)$. The reason for doing so is that $L(X)$ is built up from below in a definable fashion and the enables us to track sets via their definitions. This is in contrast to the inner model HOD. There are, however, other models which are richer than $L(X)$ and yet resemble $L(X)$ in being built up from below in a definable fashion—for example, the core model K_{DJ} constructed by Dodd and Jensen. The problem of tracking is thus solved for K_{DJ} and,

reasoning as above in the case for $L(X)$, we can conclude that conclude that there is a non-trivial elementary embedding from K_{DJ} into itself.[8] More generally let us formulate the *general extension principle* thus: Suppose that V_κ and V_λ are legitimate candidates for V. Suppose that K is an inner model which is built up from below in a definable fashion and thus admits a notion of tracking. Then there is a non-trivial elementary embedding $j : K(V_\kappa) \to K(V_\lambda)$ where $K(X)$ is constructed in the same manner as K only with X as the starting point. To make this principle precise and definite we must specify the class K of models for which the above schema hold.

The inner models that I will need for present purposes are as follows: Steel's core model $K(x)$ and the models $L_\kappa^{\mathscr{M}_n^\#}(V_\alpha)$ obtained by taking the constructible closure of V_α and the ordinals under the function $\xi \mapsto \mathscr{M}_n^\#(x)$, that is, letting F be the (class size) function taking x to $\mathscr{M}_\kappa^\#(x)$, $L_\kappa^{\mathscr{M}_n^\#}(V_\alpha) = L^F$. Both of these models are built up from below in a definable fashion and so must be *non-rigid*, that is, there must be an elementary embedding from each model into itself. Given this we can argue for the following principle:

GEP. *For each x the model $K^c(x)$ reaches a Woodin cardinal and for each α and for each n the model $L_\kappa^{\mathscr{M}_n^\#}(V_\alpha)$ is not rigid.*

The second clause is immediate. The first clause follows from the fact that K is not rigid and the following theorem of Steel, a proof of which can be found in Steel (1996) p. 88 ff. Let K be Steel's core model.

THEOREM 40. (Steel) *Suppose that K is not rigid. Then $K^c \models$ 'There is a Woodin cardinal'.*

4.4.2 Projective Determinacy. The principle GEP implies that K_{DJ} is not rigid and by a result of Dodd and Jensen this implies that there is an inner model with a measurable cardinal. This gives us the consistency but not necessarily the truth of the statement that there is a measurable cardinal. Our primary concern, however, is with truth and not consistency. In this subsection I will show that GEP implies a host of interesting things which are not provable in ZFC + EP. The central result

is the following, which is a straightforward application of Woodin's technique of the core model induction.

THEOREM 41. GEP *implies* PD.

Proof. Since, by a result of Woodin, PD is equivalent to the statement '$\forall x \in \mathbb{R} \mathscr{M}_n^{\#}(x)$ exists', it suffices to prove the latter by induction on n. The base case $n = 0$ is clear as noted in §4.3.1. For each α consider $L_\kappa^{\mathscr{M}_n^{\#}}(V_\alpha)$. By GEP we have that for arbitrarily large ν there is an amenable measure on $\mathscr{P}(\nu) \cap L_\kappa^{\mathscr{M}_n^{\#}}(V_\alpha)$. For each such ν and for each $x \in L_\kappa^{\mathscr{M}_n^{\#}}(V_\alpha)$ we can build $K^c(x)$ up to ν as in Steel (1996). Let us denote the resulting model by $K_\nu^c(x)$.

CLAIM. *For every α and for every ν as above the model $K_\nu^c(x)$ has a level which is not $n+1$-small.*

Proof. (Sketch) Suppose not. Fix α and ν as above such that all levels of $K_\nu^c(x)$ are $n+1$-small. One can show that $K_\nu^c(x) \models$'There does not exist a Woodin cardinal'. (For details see Steel (2001)). But this contradicts GEP. □

Thus for each α there is a premouse $\mathscr{N}_\alpha(x)$ such that (i) $\mathscr{N}_\alpha(x) \equiv \mathscr{M}_{n+1}^{\#}(x)$ (ii), $\mathscr{P}_\alpha$ projects to Card(x), and (iii) $\mathscr{N}_\alpha(x)$ is iterable for trees in V_α. Now suppose that α and β are greater than Card$(x)^+$. Since $\mathscr{N}_\alpha(x)$ is iterable for trees in V_α and $\mathscr{N}_\beta(x)$ is iterable for trees in V_β we can coiterate $\mathscr{N}_\alpha(x)$ and $\mathscr{N}_\beta(x)$. Since both models project to Card(x) we have that one model is an initial segment of the other (cf. Steel (2000b) Corollary 3.12). Since both models are elementarily equivalent to $\mathscr{M}_{n+1}^{\#}(x)$ this means that $\mathscr{N}_\alpha(x) = \mathscr{N}_\beta(x)$. Since α was arbitrary this means that we have a model which is elementary equivalent to $\mathscr{M}_{n+1}^{\#}$ and is fully iterable. Thus $\mathscr{M}_{n+1}^{\#}(x)$ exists. Since x was arbitrary we have shown by induction that $\mathscr{M}_n^{\#}(x)$ exists for all n and for all x. Thus PD holds. □

Recall that ZFC + EP lifted all of the results of classical descriptive set theory up to the next projective level. In particular, ZFC + EP implies that there is no paradoxical decomposition of the sphere in which the pieces are $\underset{\sim}{\Sigma}^1_2$, a statement which is independent of ZFC. Unfortunately, the statement that there is no paradoxical

decomposition of the sphere in which the pieces are $\underset{\sim}{\Sigma}^1_3$ is independent of ZFC + EP. GEP closes off this independence. In fact we have that for each $n < \omega$ there is no paradoxical decomposition of the sphere in which the pieces are $\underset{\sim}{\Sigma}^1_n$. Moreover, GEP lifts the results of classical descriptive set theory up through the projective levels.

4.4.3 Freezing Second Order Arithmetic. The above theorem shows that GEP settles many questions at the level of second order arithmetic which are not settled by ZF. In fact, there is no known example of a natural sentence φ of second order arithmetic such that (i) φ is known to be independent of ZF and (ii) it is not known whether or not φ is true. And as before there is a theorem underlying this empirical fact.

THEOREM 42. GEP *freezes second order arithmetic.*

Proof. (Sketch) Let us say that a model M is Σ^1_n-*correct* if for each $a \in H(\omega_1)^M$ we have $M \models \varphi[a]$ iff $V \models \varphi[a]$ for each Σ^1_n-formula $\varphi(v)$. By the proof of the above theorem we have that $\mathscr{M}^{\#}_n(x)$ exists for all n and for all x. Let $\mathscr{M}$ be the result of iterating the top measure of $\mathscr{M}^{\#}_n$ out of the universe. It follows by a theorem of Woodin that $\mathscr{M}$ is Σ^1_n-correct. (See Steel (1995) §4 for a proof.) Now let $\mathbb{P}$ be a partial order in V and let $G \subseteq \mathbb{P}$ be V-generic.

CLAIM. $V[G] \models$ '$\mathscr{M}^{\#}_n$ *is iterable*'. *Thus* $(\mathscr{M}^{\#}_n)^V = (\mathscr{M}^{\#}_n)^{V[G]}$.

Proof. As in the proof of Lemma 3.1 of Steel (2002). □

For $a \in H(\omega_1)$ let $\mathscr{N}(a)$ be the constructible closure of a and the ordinals under the function $x \mapsto \mathscr{M}^{\#}_n(x)$ for each n. Thus

$$\begin{aligned} V \models \varphi[a] &\leftrightarrow \mathscr{N}(a)^V \models \varphi(a) \\ &\leftrightarrow \mathscr{N}(a)^{V[G]} \models \varphi(a) \\ &\leftrightarrow V[G] \models \varphi[a] \end{aligned}$$

since $\mathscr{N}(a)$ is Σ^1_n-correct and $\mathscr{N}(a)^V = \mathscr{N}(a)^{V[G]}$. Therefore, for each n, the $\underset{\sim}{\Sigma}^1_n$-theory cannot be altered by forcing, that is, second order arithmetic is frozen. □

Chapter 5

Beyond Large Cardinals

Thus far I have argued for the principle EP and the stronger principle GEP. The principle EP overcomes the Minimal Hurdle and effects a significant reduction in incompleteness. The principle GEP yields a much greater reduction in incompleteness—in particular, it implies PD and freezes the theory of second order arithmetic.

The original justifications of these principles was based on an analysis of the concept of set. In this chapter I would like to elaborate on the kind of justification introduced in §4.3.4. This kind of justification is *a posteriori* in the original literal sense of the term.

5.1 Freezing

5.1.1 The Theory of $L(\mathbb{R})$. Recall the situation in Chapter 4. I provided two types of justification of EP. The first was based on a version of the potentialist conception of set. The second was in terms of the *inevitability* of EP: EP is inevitable in the sense that it is implied by *any* (generically invariant) theory that freezes $\underset{\sim}{\Sigma}^1_3$. The nature of this second kind of justification is quite different from what is customary in mathematics. Typically in mathematics we justify a proposition from deriving it from other *basic* propositions and when asked for a justification of the basic propositions such as the axioms of PA or ZFC we appeal to the nature of the subject matter. Our second justification of EP is of neither of these two types—it does not proceed

as a derivation from more basic propositions and it does not involve an appeal to the nature of the subject matter. Instead it is a *meta-theoretical* consideration, a consideration pertaining to theories.

Our meta-theoretical justification has the following form: A good theory of subject matter S must have property P. All good theories imply X. Therefore X.[9] This new brand of justification can be applied more widely. One way to do this is to replace $\underset{\sim}{\Sigma}^1_3$ with richer pointclasses. For example, a good theory of the reals must freeze the entire projective theory, that is, it must freeze $\underset{\sim}{\Sigma}^1_n$ for all $n < \omega$. There are such theories. For instance, the theory GEP is such a theory. But here the situation differs from that of EP in that the statement that the projective theory is frozen does not imply GEP. But there are situations for which this reverse implication holds. For example, consider the theory of $L(\mathbb{R})$. This is a transfinite extension of the projective theory—the projective theory is the theory of $L_1(\mathbb{R})$. Now a good theory must freeze the theory of $L(\mathbb{R})$. It turns out that that there are extensions of ZF that can do this. For example, suppose that there is a proper class of inaccessible. Woodin showed that 'ZFC + $\mathscr{M}_\omega$ exists' freezes the theory of $L(\mathbb{R})$. Conversely, Woodin showed that *any* theory which freezes the theory of $L(\mathbb{R})$ actually implies that $\mathscr{M}_\omega$ exists. This provides a meta-theoretical justification of the axiom '$\mathscr{M}_\omega$ exists'. It is inevitable in the sense that EP was.

5.1.2 Overlapping Consensus. Let me recast the notion of inevitability in different language. Suppose that two people A_1 and A_2 each have theories which are good in the sense that they freeze the theory of $L(\mathbb{R})$. A_1 and A_2 might have incompatible theories. For example, A_1 might believe T_1 = ZFC + PFA while A_2 might believe T_2 = ZFC + CH+'There is a proper class of Woodin cardinals'. Both T_1 and T_2 imply that the theory of $L(\mathbb{R})$ is frozen. But the theories are incompatible. Fortunately, the incompatibility has nothing to do with the success of each in freezing the theory of $L(\mathbb{R})$. The reason each theory freezes the theory of $L(\mathbb{R})$ is that each theory implies the existence of the canonical large cardinal axiom '$\mathscr{M}_\omega$ exists'. This statement lies in the overlapping consensus of T_1 and T_2 and, in fact, the entire spec-

trum of theories which freeze the theory of $L(\mathbb{R})$. In other words, there is a common core on which all good theories agree. It is this common core which we all parties have reason to believe.

5.1.3 Limitations. We have seen that EP freezes $\underset{\sim}{\Sigma}^1_3$ and that GEP freezes the entire projective theory. We have also seen that '$\mathscr{M}_\omega$ exists' freezes the entire theory of $L(\mathbb{R})$. Each of these axioms is implied by large cardinal axioms; for example, all three of these axioms are implied by the large cardinal axiom asserting that there is a proper class of Woodin cardinals. It is natural then to ask whether large cardinal axioms can freeze much more than the theory of $L(\mathbb{R})$.

In his (1946) Gödel suggested that large cardinal axioms might settle not only CH but all of the open problems of mathematics. This program—known as *Gödel's program*—has been remarkably successful, as noted above. Unfortunately, Levy & Solovay (1967) showed that none of the standard large cardinal axioms can settle CH. It follows that none of the standard large cardinal axioms can freeze the Σ^2_1-theory. Moreover, it follows from results of Woodin that large cardinals can freeze every pointclass of complexity strictly below Σ^2_1. Thus in choosing CH as a test case for his program Gödel located precisely the point at which the program fails.

5.2 Ω-Logic and the Continuum Hypothesis

5.2.1 Strong Logics. Let us recast the above results in more modern language, the language of strong logics. One of the aims of strong logics is to capture provability in ZFC + 'Large Cardinal Axioms' in a single logic. Such a logic will make precise Gödel's suggestion that large cardinal axioms yield the notion of *absolute demonstrability.* Henceforth let us assume every large cardinal axiom not known to be inconsistent.

From the model theoretic perspective first-order logic can be characterized as the set of sentences true in all structures of the form (M, E) where E is a binary relation on M. Following Woodin (1999), let Γ be a collection of *test structures* of the form

$\mathscr{M} = (M, E)$ where $E \subseteq M \times M$ and define:

$$\text{ZFC} \vdash_\Gamma \varphi \text{ iff } \forall \mathscr{M} \in \Gamma \, (\mathscr{M} \models \text{ZFC} \rightarrow \mathscr{M} \models \varphi).$$

As special cases we have first-order logic (by letting Γ consist of *all* test structures), ω-logic (by letting Γ consists of test structures in which ω is isomorphic to an initial segment of E), and β-logic (by letting Γ consists of test structures such that E is well-founded). Notice that we are conditioning logic on ZFC and that as we narrow the class of test structures the logic becomes stronger.

A natural question is whether there is a strongest logic. We have to place some restrictions on Γ if $\vdash_\Gamma$ is to count as a *logic*. For if we take $\Gamma = \{V\}$ then we have $\text{ZFC} \vdash_\Gamma \varphi$ if and only if $V \models \varphi$, and it is clear that *truth in* V is not a logic. There are minimal conditions which I think that everyone will admit. Let me describe two of these: *generic soundness* and *generic invariance*. A logic $\vdash_\Gamma$ is *generically sound* iff whenever $\text{ZFC} \vdash_\Gamma \varphi$ and $V^{\mathbb{P}}_\alpha \models \text{ZFC}$ then $V^{\mathbb{P}}_\alpha \models \varphi$. This says that you can add each $V^{\mathbb{P}}_\alpha$ to Γ without perturbing the logic. A logic $\vdash_\Gamma$ is *generically invariant* iff for each partial order $\mathbb{P}$, $V \models \text{ZFC} \vdash_\Omega \varphi$ iff $V^{\mathbb{P}} \models \text{ZFC} \vdash_\Omega \varphi$. Let us say that $\vdash_\Gamma$ is a *strong logic* iff it is generically sound and generically invariant. It is clear then that there is a strongest logic which is generically sound, namely, the logic defined by taking as test structures all and only structures of the form $V^{\mathbb{P}}_\alpha$. This logic is called Ω^*-*logic*. Furthermore, under our background assumptions, all of the examples we have considered so far are generically sound and generically invariant and are thus strong logics.

There is a natural hierarchy of increasingly strong systems of strong logics that includes first-order logic, ω-logic, β-logic and reaches much beyond to a limit called Ω-logic. The definition of Ω-logic is natural but complicated. Fortunately, the actual definition is not important for our purposes.[10] What is important is that Woodin showed that Ω-logic is indeed a strong logic, that is, it is generically sound and generically invariant. (These results are provable in the theory ZFC + 'there is a proper class of Woodin cardinals', which is much weaker than our background theory.)

A question which will play a role in the sequel is the following: Is Ω-logic as strong as Ω^*-logic (in the sense that the two logics prove the same Π_2-sentences)? The Ω-*conjecture* is the statement that this is indeed the case.

5.2.2 Gödel's Program. We are now in a position to restate the success of Gödel's program.

THEOREM 43. (Woodin) *Suppose that there is a proper class of Woodin cardinals. Then either*

(1) $\mathrm{ZFC} \vdash_\Omega$ '$H(\omega_1) \models \varphi$' *or*

(2) $\mathrm{ZFC} \vdash_\Omega$ '$H(\omega_1) \models \neg\varphi$'

for each sentence φ.

Furthermore, let φ be a sentence of complexity less than that of CH. Then either $\mathrm{ZFC} \vdash_{\Omega^*} \varphi$ or $\mathrm{ZFC} \vdash_{\Omega^*} \neg\varphi$. How much further can one go? Is *demonstrable in* Ω^**-logic* the absolute notion of demonstrability, that is, does Ω-logic settle all questions of set theory?

Interestingly, in a postscript to his 1964 paper (dated September 1966) Gödel wrote that in light of Cohen's results

> it seems to follow that the axioms of infinity mentioned on footnote 20, to the extent to which they have so far been precisely formulated, are not sufficient to answer the question of the truth or falsehood of Cantor's continuum hypothesis. (Gödel 1964, 273)

As noted above this was later proved by Levy and Solovay in their paper of 1967.[11] So one cannot settle CH in Ω^* logic (or Ω-logic) but one can settle every question of strictly less complexity than that of CH. Thus, in choosing CH as a test case for his program Gödel located the precise point at which his program fails.

One must therefore move beyond standard large cardinal axioms if one is to have a hope of settling CH and (more optimistically) isolate the concept of absolute demonstrability. In anticipation of this possibility Gödel wrote that

> there may exist, besides the usual axioms, the axioms of infinity, and the axioms mentioned in footnote 18 [higher-order reflection principles], other (hitherto unknown) axioms of set theory which a more profound understanding of the concepts underlying logic and mathematics would enable us to recognize as implied by these concepts (see, e.g. footnote 23). (Gödel 1964, 265)

Note that even at this point Gödel is still referring to the analysis of the fundamental notions of set theory. This is the most secure footing one could have. But a posteriori considerations my also (and perhaps necessarily) play a role.

> There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole field , and yielding such powerful methods for solving problems (and even solving them constructively, as far as that is possible) that, no matter whether or not they are intrinsically necessary, they would have to be accepted at least in the same sense as any well-established physical theory. (Gödel 1964, 265)

5.2.3 The Continuum Hypothesis. In footnote 23 Gödel mentions principles "which (similar to Hilbert's completeness axiom in geometry) would state some maximum property of the system of all sets" in contrast to the axiom $V = L$ which states a minimum property. He also notes that "only a maximum property would seem to harmonize with the concept of set explained in footnote 14". There are maximality principles which have implications for the size of the continuum. For example, the Proper Forcing Axiom and Martin's Maximum imply that the size of the continuum is $\aleph_2$. The most fruitful maximality axiom is Woodin's axiom $(*)$. This axiom implies that the size of the continuum is $\aleph_2$ but it does much more.

THEOREM 44. (Woodin) *Suppose that there is a proper class of Woodin cardinals. Then either*

(1) $\mathrm{ZFC} + (*) \vdash_\Omega$ '$H(\omega_2) \models \varphi$' *or*

(2) $\mathrm{ZFC} + (*) \vdash_\Omega$ '$H(\omega_2) \models \neg\varphi$'

for each sentence φ.

Thus, in the context of Ω-logic the theory of $H(\omega_2)$ is finitely axiomatisable via $(*)$ over ZFC. What is amazing is that through a long series of theorems Woodin managed to show that such a situation cannot happen unless CH *fails.*

THEOREM 45. (Woodin) *Suppose that there is a proper class of Woodin cardinals. Let* ψ *be a sentence such that there is a cardinal* κ *such that* $V_\kappa \models \mathrm{ZFC} + \psi$. *Suppose that either*

(1) $\mathrm{ZFC} + \psi \vdash_\Omega$ '$H(\omega_2) \models \varphi$' *or*

(2) $\mathrm{ZFC} + \psi \vdash_\Omega$ '$H(\omega_2) \models \neg\varphi$'

for each sentence φ. *Then* CH *is false.*

This provides an argument against CH. Let us say that an axiomatization $\mathrm{ZFC}+A$ of the structure $H(\omega_2)$ is *good* if A is finite and $\mathrm{ZFC}+A$ is a complete axiomatization of $H(\omega_2)$ in Ω-logic. Thus, we have a good theory of $H(\omega_2)$, namely, $\mathrm{ZFC} + (*)$. Furthermore, this implies that CH is false (in fact, it implies that the size of the continuum is $\aleph_2$). Even more, we know that the *only* way we can have a good theory of $H(\omega_2)$ is if CH is false. On considerations of simplicity one could argue that this is reason to believe that CH is false. To strengthen the case one would have to show that Ω-logic is indeed the strongest logic, that is, one would have to prove the Ω-conjecture.

(One thing that is interesting about the above argument is that it is based on considerations of simplicity. Some philosophers of science have maintained that if one has to choose between two physical theories that agree on a large domain (that includes the empirical data and possibly other further essential principles), then one should choose the simpler theory. Even more, it is often maintained that the simple theory is closer to the truth. But in physics such situations do not arise naturally.[12] Here we have such a situation arising at the forefront of mathematical research. What is the

analogue of the empirical data? One could argue that it contains every mathematical proposition of set theory which is implicit in the iterative conception of set theory, say ZF and the small large cardinal axioms and that beyond that one is guided by considerations of simplicity and the like.)

5.3 The Structure Theory of $L(V_{\lambda+1})$

5.3.1 Analogy. There is another approach to resolving CH which is worth mentioning. This approach is also due to Woodin. We saw above that large cardinal axioms freeze the theory of second-order arithmetic and much more. In particular, they freeze the theory of the structure $L(\mathbb{R})$. The key to the structure theory of $L(\mathbb{R})$ is contained in a single axiom, the axiom $\mathrm{AD}^{L(\mathbb{R})}$ stating that every set of reals in $L(\mathbb{R})$ is determined. Now at the upper end of the large cardinal spectrum there is an axiom which asserts that there is a non-trivial elementary embedding $j : L(V_{\lambda+1}) \to L(V_{\lambda+1})$ with critical point less than λ. What is amazing is the parallel between $L(\mathbb{R})$ under the axiom $\mathrm{AD}^{L(\mathbb{R})}$ and $L(V_{\lambda+1})$ under the above embedding axiom. Here λ is the analogue of ω and $V_{\lambda+1}$ is the analogue of $\mathbb{R}$. As an example of the parallel we note that under the respective axioms ω_1 is measurable in $L(\mathbb{R})$ and λ^+ is measurable in $L(V_{\lambda+1})$. This parallel is just the tip of the iceberg. There is a key difference though: whereas it is clear that $\mathrm{AD}^{L(\mathbb{R})}$ is the entire story as far as $L(\mathbb{R})$ is concerned it is *not* clear that the above embedding axiom is the entire story as far as $L(V_{\lambda+1})$ is concerned. It is very likely that structure theory of $L(V_{\lambda+1})$ will require more than the embedding axiom. What has this got to do with CH? Well, it is possible that in pinning down the structure theory of $L(V_{\lambda+1})$ we will settle CH. The reason is that the structure theory of $L(V_{\lambda+1})$ may involve axioms which (unlike the standard large cardinal axioms) interfere with the forcing machinery for proving the independence of CH. Woodin has already pointed out a possible example: Let

$$S^{\lambda^+}_{\delta} = \{\alpha < \lambda^+ \mid \mathrm{cof}(\alpha) = \delta\}$$

where δ is an infinite regular cardinal less than λ^+. Let $\mathscr{I}^{\lambda^+}_{\mathrm{NS}}$ be the ideal of nonstationary subsets of λ^+. The following is (arguably) a plausible axiom for the structure theory of $L(V_{\lambda+1})$:

AXIOM. *For each infinite regular cardinal $\delta < \lambda^+$*

$$\left(\mathscr{P}(S^{\lambda^+}_\delta)/\mathscr{I}^{\lambda^+}_{\mathrm{NS}}\right)^{L(V_{\lambda+1})}$$

is trivial.

Assume the axiom. The point is that if CH fails then the standard way of forcing it to be true kills the Axiom and if CH holds then it is not clear that one can force it to fail without killing the Axiom. So the following question is open:

QUESTION. *Does the above axiom settle* CH*?*

5.3.2 Conclusion. In this chapter I have discussed two types of justification in mathematics which are quite different that the kinds of justification which are customary in mathematics. The first kind of justification is *meta-theoretic*. This kind of justification has the form: Every "good" theory implies X. We saw two examples of this form of justification. First, in §5.1 the good theories were those which implied that a given pointclass—such as $\underset{\sim}{\Sigma}^1_3$ or the theory of $L(\mathbb{R})$—is frozen. We saw that all good theories in this sense have strong implications. For example, in the case of $\underset{\sim}{\Sigma}^1_3$ all good theories imply SPHERE; in the case of $L(\mathbb{R})$ all good theories imply $\mathrm{AD}^{L(\mathbb{R})}$. Second, in §5.2 the good theories were those that provided a finite axiomatisation of $H(\omega_2)$ in the context of Ω-logic. We saw that all good theories in this sense have strong implications. For example, all such good theories imply that CH is false. The second kind of justification is *analogical*. In §5.3 we saw that principles of analogy bridging $L(\mathbb{R})$ and $L(V_{\lambda+1})$ may have implications for the size of the continuum.

Notes

[1]This section was inspired by a claim of Parsons (cited by Feferman in his introduction to Gödel (*1933o)) to the effect that the approach of Gödel (*1933o) does not yield a justification of Replacement. Tait (2001) argues that it does. This conflict let me to undertake a close examination of Gödel's texts and reconstruction of Gödel's argument.

[2]Notice that Gödel is taking the powerset operation as primitive. In other places (such as Gödel (1964)) he takes the operation 'set of' as primitive. He seems to regard the latter as more primitive than the former (see, for example, his statements in Wang (1996, 8.2.13, 8.2.17), but in his *1951 he writes: "The operation 'set of' is substantially the same as the operation 'power set'..." (fn. 5). See Parsons (1995), 86–88 for an illuminating discussion of the powerset operation.

[3]One thing that I find amazing about this quotation (and the end of the *1951 passage quoted above) is that Gödel seems to be espousing the view that the universe of sets is a *potential*, not *actual*, totality. This is in sharp contrast to the common view that Gödel thought of the universe as closed. (There are similar remarks to be found in Cantor who, like Gödel, is taken as classic example of one who held that the universe of sets is a completed totality (in contrast to, say, Zermelo). For example, Parsons (1974) writes "Cantor, in 1899, distinguished what he called "inconsistent multiplicities" (*inkinsistente Vielheiten*) by appealing to the irreducibly potential character of the "totality" of sets; the latter are such that a contradiction results from supposing a *Zusammensein* of all their elements (GA, p. 443)".)

[4]This borders on being a tautology. It seems that if one believes T then one must believe Con(T). Of course, it is *consistent* to believe T + ¬Con(T) but it is not *coherent* to believe it. A possible response to this last claim is the following, something we might call the *sorites* response: As one iterates the addition of the consistency statement the consistency strength of the successive systems increases

and so the likelihood that the system is inconsistent increases; but it is perfectly reasonable to have doubts about the consistency of some of these higher systems and so at some point in the procession it must be reasonable to believe $T + \neg Con(T)$. This response gets its force from the idea that at some point in the progression it is reasonable to doubt the consistency of the system. But it is crucial here to be specific about what one means by "the progression". If by "the progression" one means the ω-length progression then it seems to me that it *is* incoherent to maintain $T + \neg Con(T)$ for any theory in the progression. There is no loss in confidence from one level to the next. If, in contrast, one intends "the progression" to include a much longer progression then I agree that there may be points at which one's doubts reasonably arise but I believe that they arise for a different reason. For example, suppose that one is a staunch predicativist. Then one will accept the iteration of consistency sentences (or local reflection principles) up to any ordinal $\alpha < \Gamma_0$. And, letting T be the base theory and T_α the α^{th}-theory in the progression, one will accept $T_\alpha + Con(T_\alpha)$. But doubts will arise at the ${\Gamma_0}^{th}$-stage, not because one accepts T_{Γ_0} but not $Con(T_{\Gamma_0})$, but rather because, granting the predicativist view, one cannot even make sense of T_{Γ_0}—predicative analysis is not predicatively characterisable. In short, doubts arise not at successor steps of the progression but rather at those limit stages where one has doubts about the mathematical machinery required to make sense of the stage.

[5]Two points are in order. First, although the actualist cannot express the claim that the height of the universe is weakly compact there may be a way for the actualist to motivate the claim that there is a *level* of the universe the height of which is weakly compact. Second, some actualists might try to maintain that full second-order quantification over V is available. For example, the advocate of the plural quantificational interpretation of second order languages might maintain this. I have not seen a convincing argument for such a position. I won't argue the issue here. It suffices to note that even if we grant the advocate of plural quantification free access to second order logic he will have trouble with third or higher order logics. This

suffices to make my point that the actualist is limited in a way that the potentialist is not.

[6]It is an interesting open question whether the inconsistency requires AC. In fact, there are vast strengthenings of Reinhardt's axiom not known to be inconsistent without assuming AC. There is an entire hierarchy of "choiceless cardinals" and it may be the case that the hierarchy of consistency strength outstrips that which assumes choice. In the end it may turn out to be reasonable to view AC as a limitative axiom on a par with $V = L$.

[7]I am indebted to Richard Ketchersid for very useful discussion of this and the next section.

[8]Dodd and Jensen showed that this is equivalent to the statement that there is an inner model with a measurable cardinal. So we have a justification of such a model. Note, however, that this is quite different from a justification of the existence of a measurable cardinal. A further argument would be required to move from the consistency to the existence of a measurable cardinal. I suspect that such an argument can be supplied—large cardinals (in contrast, say, to an ω_2-well-ordering of the reals) seem to be the type of things which require for their existence only their consistency. But I will not pursue this thought here.

[9]There is a natural objection to this form of argument, namely, that there might not be any good theories. Let us call this the *wishful thinking objection.* The objector acknowledges that it would be nice if we had a good theory but maintains that it is just wishful thinking to maintain that there are any good theories. This objection is best met on a case by case basis. For the moment let me focus on EP.

In the case of EP the objector maintains that it is mere wishful thinking to think that there is theory which freezes $\underset{\sim}{\Sigma}^1_3$. Let me isolate a strong form of this objection. According to the strong form, the objector is maintaining that it is wishful thinking to think that there is a good theory which is *consistent.* Of course, the objector

need not make such a strong claim; rather, the objector need only maintain that it is wishful thinking to think that there is a good theory which is *true*.

My reason for isolating the above strong version of the objection is that I want to argue against it as a warm-up to an argument against the weaker objection. I think that we can provide good reasons for the consistency of EP. These reasons are tied up with inner model theory: If EP is inconsistent then we can derive a contradiction by assuming it. So assume EP. It follows from inner model theory that there are fine structural models of EP. But these models involve a lot of control—they are much like Gödel's L, in contrast to HOD or V. If there were a contradiction lurking in the background then it is likely that it would be brought to light in the context of a fine structural model. Thus there is good reason for believing in the consistency of EP.

Let us then assume that the objector admits that there is good reason to believe that EP is consistent but still maintains that it is mere wishful thinking to think that EP is true. There are two steps in my response to this. First, we have seen that EP is equivalent to 'for all X, $X^{\#}$ exists'. So the objector must admit that there are models of 'for all X, $X^{\#}$ exists'. Second, the statement that $X^{\#}$ exists is absolute for proper class transitive $\in$-models of ZF containing X. Thus if there is a proper class transitive $\in$-model of ZF satisfying 'for all X, $X^{\#}$ exists' then this statement is actually true. It follows that the objector must maintain (i) that there are models of 'ZF + For all X, $X^{\#}$ exists' but that (ii) there are no proper class transitive $\in$-models of 'ZF + For all X, $X^{\#}$ exists'. Now this is a logically consistent position but it is one which is difficult to maintain given the structuralist view of mathematical objects. It is analogous to maintaining that PA is consistent but that there are no wellfounded models of PA. This too is a logically consistent position but it is hard to maintain since from any non-wellfounded model of PA we can extract a wellfounded model of PA. In short, it is hard to maintain that there are *only* non-standard models of a given theory, especially given that there is a canonical method for isolating standard models from a given non-standard model. I think that this line of argument goes some way in meeting the concerns of the objector.

[10]For those who are interested I will give the definition of Ω-logic. First, I need to introduce the notion of a *universally Baire* set of reals and the notion of a structure being *A-closed* where A is a universally Baire set of reals. A set $A \subseteq \mathbb{R}$ is *universally Baire* iff there are trees T and T^* such that $A = p[T]$, $\mathbb{R} - A = p[T^*]$, and, for every partial order $\mathbb{P}$, $V^{\mathbb{P}} \models p[T] \cup p[T^*] = \mathbb{R}$. That is, A and its complement $\mathbb{R} - A$ have tree representations which continue to project to complements in generic extensions. So if $A \subseteq \mathbb{R}$ is universally Baire and $V[G]$ is a generic extension, there is a canonical representation $A_G \subseteq \mathbb{R}^{V[G]}$ in $V[G]$, namely, $p[T]^{V[G]}$. Universally Baire sets are thus *definable* in a rich sense. Now suppose $A \subseteq \mathbb{R}$ is universally Baire and M is a transitive model of ZFC. Then M is *A-closed* iff for each partial order $\mathbb{P} \in M$ and V-generic $G \subseteq \mathbb{P}$ we have $V[G] \models A_G \cap M[G] \in M[G]$. Let Γ^∞ be the set of all universally Baire sets. We can now define Ω-logic: Suppose there is a proper class of Woodin cardinals. Then ZFC $\vdash_\Omega \varphi$ if there is a $A \in \Gamma^\infty$ such that φ is true in every M which is a countable transitive model of ZFC that is A-closed. For $\Gamma \subseteq \Gamma^\infty$, Ω_Γ-logic is defined similarly, with Γ in place of Γ^∞. The resulting directed system $\langle \Omega_\Gamma \mid \Gamma \subseteq \Gamma^\infty \rangle$ of logics includes ω-logic, β-logic and has Ω-logic at the top.

[11]Did Gödel know of this result? If so then why the caution? Perhaps he anticipated the proof.

[12]The only examples I can think of are the Bohmian and the Ghirardi-Rimini-Weber formulations of quantum mechanics. Both of these formulations are (as far as is known) empirically equivalent to the standard formulation of quantum mechanics. Yet considerations of simplicity (largely ontological) and/or clarity (in avoiding the vague notion of "measurement") have some—mainly philosophers—to prefer one of these two formulations to the standard formulation.

Bibliography

Ackermann, W. (1956). Zur Axiomatic der Mengenlehre, *Mathematische Annalen* (141): 336–45.

Feferman, S. (1996). Gödel's program for new axioms: Why, where, how and what?, *in* P. Hajek (ed.), *Godel 96: Logical Foundations of Mathematics, Computer Science, and Physics*, number 6 in *Lecture Notes in Logic*, AK Peters Ltd., pp. 3–22.

Gödel, K. (*1933o). The present situation in the foundations of mathematics, *in Gödel (1995)*, Oxford University Press, pp. 45–53.

Gödel, K. (1946). Remarks before the princeton bicentennial conference on problems in mathematics, *in Gödel (1990)*, Oxford University Press, pp. 150–153.

Gödel, K. (*1951). Some basic theorems on the foundations of mathematics and their implications, *in Gödel (1995)*, Oxford University Press, pp. 304–323.

Gödel, K. (1964). What is Cantor's continuum problem?, *in Gödel (1990)*, Oxford University Press, pp. 264–270.

Gödel, K. (1990). *Collected Works, Volume II: Publications 1938–1974*, Oxford University Press, New York and Oxford.

Gödel, K. (1995). *Collected Works, Volume III: Unpublished Essays and Lectures*, Oxford University Press, New York and Oxford.

Kanamori, A. (1997). *The Higher Infinite*, Perspectives in Mathematical Logic, Springer-Verlag, Berlin.

Levy, A. & Solovay, R. M. (1967). Measurable cardinals and the continuum hypothesis, *Israel Journal of Mathematics* **5**: 234–248.

Lindström, P. (1997). *Aspects of Incompleteness*, Springer-Verlag.

Martin, D. A. & Steel, J. R. (1989). A proof of projective determinacy, *Journal of the American Mathematical Society* **2**(1): 71–125.

Parsons, C. (1974). Sets and classes, *in Parsons (1983)*, Cornell University Press, pp. 209–220.

Parsons, C. (1977). What is the iterative conception of set?, *in Parsons (1983)*, Cornell University Press, pp. 268–297.

Parsons, C. (1983). *Mathematics in Philosophy: Selected Essays*, Cornell University Press, Ithaca, New York.

Parsons, C. (1995). Structuralism and the concept of set, *in* W. S.-A. et al. (ed.), *Modality, Morality, and Belief: Essays in honor of Ruth Barcan Marcus*, Cambridge University Press.

Quine, W. V. O. (1969). *Set Theory and Its Logic*, second edn, Harvard University Press, Cambridge, Massachusetts.

Quine, W. V. O. (1986). *Philosophy of Logic*, second edn, Harvard University Press, Cambridge, Massachusetts.

Reinhardt, W. (1967). *Topics in the Metamathematics of Set Theory*, PhD thesis, University of California, Berkeley.

Reinhardt, W. (1968). Some strong axioms of infinity. Manuscript.

Reinhardt, W. (1974). Remarks on reflection principles, large cardinals, and elementary embeddings, *Proceedings of Symposia in Pure Mathematics*, Vol. 10, pp. 189–205.

Schimmerling, E. (2001). The ABCs of mice, *Bulletin of Symbolic Logic* **7**(4): 485–503.

Shoenfield, J. (1967). *Mathematical Logic*, Addison-Wesley Series in Logic, Addison-Wesley Pub. Co., Reading, Massachusetts.

Steel, J. (1995). Projectively well-ordered inner models, *Annals of Pure and Applied Logic* (74): 77–104.

Steel, J. (1996). *The Core Model Iterability Problem*, Lecture Notes in Logic.

Steel, J. (1999). Lectures on inner model theory. Berkeley.

Steel, J. (2000a). Mathematics needs new axioms, *Bulletin of Symbolic Logic* **6**(4): 422–433.

Steel, J. (2000b). Outline of inner model theory. To appear in the *Handbook of Set Theory.*

Steel, J. (2001). Some notes on the core model induction. Handwritten notes taken by Philip Welch.

Steel, J. (2002). Core models with more Woodin cardinals, *Journal of Symbolic Logic* **67**: 1197–1226.

Tait, W. W. (1990). The iterative hierarchy of sets, *Iyyun* **39**: 65–79.

Tait, W. W. (1998a). Foundations of set theory, *in* H. Dales & O. G. (eds), *Truth in Mathematics*, Oxford University Press, pp. 273–290.

Tait, W. W. (1998b). Zermelo on the concept of set and reflection principles, *in* M. Schirn (ed.), *Philosophy of Mathematics Today*, Oxford: Clarendon Press, pp. 469–483.

Tait, W. W. (2001). Gödel's unpublished papers on foundations of mathematics, *Philosophia Mathematica* **9**: 87–126.

Wang, H. (1974). *From Mathematics to Philosophy*, Routledge & Kegan Paul, London.

Wang, H. (1977). Large sets, *in* Butts & Hintikka (eds), *Logic, Foundations of Mathematics, and Computability Theory*, D. Reidel Publishing Company, Dordrecht-Holland, pp. 309–333.

Wang, H. (1996). *A Logical Journey: From Gödel to Philosophy*, MIT Press.

Woodin, W. H. (1982). On the consistency strength of projective uniformization, *in* J. Stern (ed.), *Proceedings of the Herbrand Symposium. Logic Colloquium '81*, North-Holland, Amsterdam, pp. 365–383.

Woodin, W. H. (1999). *The Axiom of Determinacy, Forcing Axioms, and the Nonstationary Ideal*, Vol. 1 of *de Gruyter Series in Logic and its Applications*, de Gruyter, Berlin.