A Logical Foundation For Potentialist Set Theory

Sharon Berry

September 16, 2020

Contents

Part I

Chapter 1

Introduction

"Admittedly, the present state of affairs where we run up against the paradoxes is intolerable. Just think, the definitions and deductive methods which everyone learns, teaches, and uses in mathematics, the paragon of truth and certitude, lead to absurdities! If mathematical thinking is defective, where are we to find truth and certitude?[1] *"*

– David Hilbert, *On the Infinite*

[1]translation from [4]

1.1 Mathematics as a Touchstone and the Centrality of Set Theory

Mathematical proofs provide a touchstone of clarity and convincingness which serves as an inspiration to philosophy and other disciplines. While it is possible to doubt the results of mainstream mathematical arguments (philosophers are capable of doubting anything), there's something striking about just how convincing mathematical proofs often are. Consider the standard argument that there are infinitely many primes. Even philosophers who deny that there are numbers (and hence think the argument as usually stated is unsound) are strongly tempted to say that we know *something like* the premises and that the proofs provide some kind of valuable amplification of this knowledge. The premises we use in informal mathematical reasoning have a combination of prima facie obviousness and power/generality, which makes them exemplary tools for expanding our knowledge and resolving disputes in cases where people's initial hunches disagree. It's no surprise that Leibniz[2] wished philosophers could resolve their disputes like mathematicians by saying 'let us calculate' (or at least, 'let us each look for a proof').

Now (in many ways) set theory lies at the heart of modern mathematics, and it does powerful mathematical (not just philosophical) work as a foundation for the whole. So one might hope that the set theoretic foundations for mathematics would share the clarity and convincingness we hope for from mathematical arguments.

[2] See page 14 of [21].

However, certain problems in the philosophical foundations of set theory raise serious questions and doubts about our acceptance of the axioms of set theory. These questions are more subject matter specific and threatening to things normal mathematicians care about than generic philosophical doubts about whether there are any abstract objects, or whether the knowledge we get from the standard proof that there are infinitely many primes is better construed nominalistically or Platonistically.

Specifically, the development of set theory resolved a great many problems in analysis. It also provided a formal framework to allow interactions between various areas of mathematics creating, as Hilbert famously observed [58], a kind of mathematical paradise. However contradiction threatened Hilbert's paradise, in the form of Russell's paradox.

This problem was almost but not quite solved by accepting the iterative hierarchy conception of sets and the standard Zermelo-Fraenkel with choice (ZFC) first-order axioms for set theory. On the iterative hierarchy conception of set we think about the sets as being formed in layers, with the sets at each layer containing only elements from prior layers. So (departing slightly from the mathematical approach which identifies stages with sets) we can think of the hierarchy of sets as a two sorted structure consisting of:

- a well ordered series of stages and
- sets formed at these stages, such that a set is formed at a stage iff its members are all formed at earlier stages.

And now set theory is widely accepted as a foundation for all of modern mathematics. It is hard to deny that the mathematical results which are currently stated in terms of set theory reflect genuine and important knowledge of some kind. But a question of **how to justify these axioms** remains.

So we may ask: is the price of remaining in Cantor's set theoretic paradise giving up the old ambition of founding mathematics on intrinsically obvious seeming principles?

One of the main projects of this book will be to develop a unified determinate conception of set theoretic truth, which vindicates our intuitive expectations regarding set theory – and makes all theorems of ZFC set theory derivable from (prima facie) obvious premises as traditionally desired in mathematics.

My proposal attempts to improve on standard 'actualist' approaches to set theory, which fall short of this ideal in several ways.

1.2 Actualism and its Discontents

According to standard actualist approaches to set theory, set theory is about objects called 'sets', which exist outside of space and time. On this view, sets are abstract mathematical objects, just like the natural numbers (on a Platonist understanding of the natural numbers). Apparent existence claims made by set theorists (like 'there is a set which has no elements') are made true by the existence of corresponding objects, just like ordinary existence claims about cities or electrons or cars.

Actualists run into three problems (each of which I'll develop in much greater detail in Chapter 2). First, actualist approaches don't offer a determinate conception of set theoretic structure. In particular, the height of the hierarchy of sets is left vague or mysterious. The Burali-Forti paradox[3] dramatizes, it appears that once that a certain naive conception of the hierarchy of sets (as containing ordinals corresponding to all ways some objects are well ordered by some relation) is incoherent. And once this naive conception is rejected it appears that, for any height that the hierarchy of sets could achieve, there could be a strictly larger structure which adds an extra layer of 'fake sets' on top of the original hierarchy, and fits with everything in our conception of the sets equally well. But it seems arbitrary to suppose that the hierarchy of sets just happens stops somewhere.

Second, as a foundation for mathematics, one might hope that set theory should be able to represent any mathematical structure one might want to study. And the idea that set theory has this kind of generality is prima facie quite intuitive. But actualist set theory is prima facie unable to represent the study of mathematical structures that are 'too large' in this sense. So actualism makes it hard to capture the intuition that 'any possible structure' should, in some sense, be fair game for treatment within set theory.

Third, actualists face a problem (which is prima facie not limited to them) of intuitively justifying certain axioms of set theory to a sufficient degree. We normally hope that the mathematical axioms which are taken as starting points for proofs will be extremely prima facie plausible (if not completely indubitable or impossible to empirically cast doubt upon). So one might hope that (once we understand set theory aright) every claim provable from

[3]See section 2.2 for more details about this.

the ZFC axioms for set theory can be shown to follow from principles that seem defeasibly obvious in this way.

However, philosophers have had significant difficulty in finding any such justification for some of the ZFC axioms of set theory. For example, in [94], Hilary Putnam writes "Quite frankly, I see no intuitive basis at all for . . . the axiom of replacement. Better put, I do not see that a notion of set on which that axiom is clearly true has ever been explained." Instead philosophers of mathematics and mathematicians have made do with less ambitious approaches to justification. For example, some have invoked external justifications, via things like the failure of mathematicians to discover any contradictions during over a century of work with ZFC. Others have shown that the axiom of replacement follows from certain powerful and plausible (but not intuitively clearly true) principles that also imply many of the other axioms of set theory. But this state of affairs can feel unsatisfying.

1.3 Potentialism and the Justification Problem

In response to the first two problems above, philosophers like Putnam, Parsons, Hellman and Linnebo[91, 55, 75, 84] have proposed that we should reject[4] actualism about set theory in favor of a different approach to set theory: potentialism. The key idea behind potentialism is that, rather than taking set theory to be the study of a single hierarchy of sets which stops at some particular point, we should instead understand set theory in terms of

[4] Strictly speaking Putnam proposes actualism and potentialism are (in some sense), two perspectives on the same thing.

modal claims about what hierarchy-of-sets-like structures are possible and how such structures could possibly be extended (I discuss potentialism in detail in chapter 3).

Merely accepting potentialism solves the first problem for actualism above. For the potentialist avoids postulating an arbitrary (or indeterminate) height for the hierarchy of sets[5]. And potentialism also plausibly solves the second problem above, by honoring the intuition that any possible mathematical structure can be studied within set theory.

However, the final problem blocking our foundational ambitions remains. Merely adopting potentialism doesn't (prima face) ensure that all theorems of mainstream set theory can be derived from premises that are clearly true. Contemporary potentialists can, and generally do, prove that (the potentailist translations of) every theorem of ZFC can be derived from *certain modal logical assumptions*. However these proofs all use principles of modal logic that aren't (and aren't claimed to) be clearly true in the way invoked by Putnam.

Existing potentialist literature has shown that potentialism is *no worse off than actualism* with regard to our foundational project and the problem of justifying replacement that Putnam raises[6]. However, potentialists have not

[5] At least potentialists like Hellman, Linnebo and Studd do avoid positing such an arbitrary stopping point for the sets. Putnam's view, on which actualist set theory and potentialist set theory are (somehow) two perspectives on the same thing, does not let us avoid this problem in any obvious way.

[6] Existing potentialists have generally adopted some version of a potentialist translation of replacement as an axiom (schema). For while these potentialist translations are not clearly true, they are (we will see) as attractive as corresponding instances of the replacement schema understood actualistically.

attempted to achieve the bold foundational project (providing a foundation for set theory in principles that seem clearly true) currently at issue.

1.4 Structure Preserving Potentialism

In this book, I will attempt to complete the foundational project evoked above by developing a modal logic (the logic of structure preservation) which is suitable to the task, and then show that this modal logic is useful for many other philosophical projects as well.

In Part I I will argue that we should indeed be potentialists about set theory for essentially the reasons indicated above, and then review major existing formulations of potentialism about set theory and some problems for each. I'll argue that these existing frameworks' choice of modal logic creates certain problems for our foundational project. So I'll propose my own formulation of potentialist set theory aimed at avoiding these problems.

A key idea will be to use a background modal logic which replace claims about what's logically possible *for objects* (as per standard quantified modal logic) with claims about what's logically possible *given certain structural facts*, expressed using a new piece of logical vocabulary I'll call the conditional logical possibility operator ($\Diamond_{...}$). Cashing out potentialist set theory in these terms lets us disentangle clearly meaningful claims in potentialist set theory from philosophical controversies about things like quantified modal logic or which concepts are indefinitely extensible (unlike prior views). In Part II I will turn to the core mathematical problem addressed in this book:

justifying the ZFC axioms from seemingly obvious premises. I'll show how to potentialistically translate set theoretic claims into the modal language I favor. I'll propose axioms for reasoning about conditional logical possibility which (I claim!) seem clearly true in the way our foundational project requires. Then I will show that potentialist translations of all theorems of ZFC set theory can be proved from these sufficiently compelling premises[7].

Finally in Part III, I will develop the story in Part I and Part II into a more general set-theoretic paradox driven, modality-first, modestly neo-Carnapian approach to mathematics as a whole. I will also argue that working with the concept of conditional logical possibility developed in Parts I and II can illuminate other philosophical topics like: grounding, neo-Carnapian theories of ontological knowledge by convention, varieties of (post) Quinean indispensability arguments, and the heterogeneity of applied mathematics.

1.5 Outline

My plan of action will be as follows.

In chapter 2 I will introduce the standard actualist approach to set theory, and note how it faces an arbitrariness problem (which is highlighted by the Burali-Forti paradox), as well as a problem about justifying the axiom of replacement.

[7] I'll show that if $ZFC \vdash \phi$ then $\phi^\Diamond$ from the intuitively obvious seeming modal principles listed before. That is, reasoning in ZFC as if talking about an actualist hierarchy of sets is harmless; if one can prove that ϕ in ZFC then the potentialist translation of ϕ (written $\phi^\Diamond$ above) is true.

In chapter 3 I'll note how adopting some form of potentialism would let us solve the above arbitrariness problem and review existing forms of the 'Putnamian' style of potentialism I favor: Putnam[95]'s original proposal and Hellman[55]'s development of it. I will defend Hellman's use of a notion of logical possibility, but note that controversies over quantified modal logic raise some problems for using his version of potentialism in our foundational project.

In chapter 4 I'll introduce my preferred style of potentialist paraphrase.and the key notion of conditional (structure-preserving) logical possibility $\Diamond_{...}$. Using this notion to formulate potentialist set theory lets us 'route around' infamous controversies concerning quantified modal logic – and remove obscure and intuitively irrelevant philosophical claims from our foundations of mathematics. Finally, in Chapter 5 I'll contrast the above approach to potentialist set theory with those advocated by Linnebo and Studd, major proponents of an alternate 'Parsonian' school of potentialist set theory.

In Part II I turn to my main mathematical project: justifying the ZFC axioms (and replacement in particular) from general principles of modal reasoning with the right kind of intuitive obviousness. I'll present my preferred formulation of potentialist set theory in the language of conditional logical possibility in **??**. Then I'll introduce a formal system for reasoning about conditional logical possibility, whose axioms I take to have the kind of prima facie obviousness needed for the current foundational project in chapters 8 through 8.

I'll prove the key theorem that (potentialist translations of) all theorems of

ZFC set theory can be reconstructed in this system in chapters B and 10 A key idea will be to use certain non-interference intuitions to justify the (potentialist translation of) the axiom of replacement, rather than simply taking the latter as an axiom, as current potentialists tend to do.

Finally in Part III of the book, I'll return to philosophical questions. In chapter 11 I'll argue that the case for potentialist set theory made in Parts I and II should be taken seriously even by philosophers with strong naturalist inclinations (despite a worry suggested by Burgess and Rosen's famous dilemma for mathematical nominalists in [18]).

I'll then consider two ways my story about set theory can be fit into a larger philosophical picture of mathematics and its applications: a nominalist approach and the weakly neo-Carnapian approach I ultimately favor.

In Chapter 16 I'll set up a key piece of machinery for both the nominalist and neo-Carnapian approach. I'll discuss a natural way of using the conditional logical possibility operator to simulate 'talking in terms of extra objects' under certain circumstances. I'll suggest this lets us do (much of) the work of Linnebo and Studd's interpretational possibility operators in more flexible and philosophically cautious way.

In Chapter 12 I'll discuss the nominalist approach to non-set theoretic mathematical objects. I'll argue that adding some cheap tricks to the above paraphrase strategy lets the nominalist answer certain classic indispensability arguments. However, I'll suggest that the mathematical nominalist *may* face serious and under-discussed worries about reference and grounding.

In Chapters 18 and 19 I'll explain the weakly neo-Carnapian approach to non-set theoretic mathematical objects I favor, and argue that adopting it helps solve the reference and grounding problems (while retaining certain benefits of nominalism). The resulting view is a kind of neo-Carnapianism realism about mathematical objects which drops Carnap's radical anti-metaphysical ambitions but keeps mathematicians' freedom to talk in terms of arbitrary logically coherent pure mathematical structures.

Finally, in chapters 20 and 22 I'll discuss how the overall picture of mathematics developed in this book relates to traditional questions about logicism, structuralism and human access to facts about objective proof-transcendent mathematical facts.

1.6 Caveats and Clarifications

Let me finish this introduction with some quick caveats about the nature and aim of my project.

First, I don't claim set theorists should literally rewrite set-theory textbooks in potentialist terms. Mathematicians' current practice of (making arguments which can be reconstructed as) proving things in first-order logic from the ZFC axioms is fine. And doing something like logical deduction from purely first-order axioms may be unavoidably easier (for minds like ours) than thinking about elaborate modal extendability claims. If one thinks about apparent first-order claims in mathematics as abbreviating potentialist claims, then the main result of Part II shows that it's unnecessary to unpack this abbreviation in mathematical contexts. For the ZFC axioms and everything derivable from them must also be true (and indeed provable from the seemingly obvious) on a potentialist reading.

However, I *am* suggesting potentialist paraphrases are what people should mean when they do and think about set theory in many philosophical contexts. They should replace current set theory with potentialist paraphrases, in these contexts because understanding set theory potentialistically solves various intuitive puzzles, and makes sense of things that we normally want to say about set theory[8].

One can think of my project of developing potentialist foundations for set theory as analogous to the familiar project of providing a set theoretic foundation for analysis. Our naive reasoning about certain concepts (limits in one case, the height of the hierarchy of sets in the other) turned out to lead to paradox. So it is desirable to find a different way of thinking about the relevant mathematical concepts which will let us capture the intuitive mathematical significance and interest of relevant mathematical claims while blocking paradoxical inferences. And it is desirable to cash out old mathematical concepts, which paradoxes may have led us to doubt that we have a coherent grip on, in other terms which we seem to understand in a way that does not invite paradox. For instance, I argue that if we cash out standard set theory in potentialist terms, the Burali-Forti paradox does not arise and yet all of mathematicians' ordinary reasoning about set theory is justified[9].

[8]See chapter 11 for much more on this point.

[9]Since I choose axioms of reasoning about conditional logical possibility which are attractive for general use (rather ones that directly mirror actualist ZF set theory), I don't obviously have (or prove) the reverse direction of the conditional '$ZFC \vdash \phi$ then $\phi^{\Diamond}$' there

Second, the potentialist understanding of pure set theory advocated in Parts I and II of this book is compatible with a range of different views about how to understand other areas of mathematics. I hope my version of potentialism will be compelling even to readers who find both nominalism and the neo-Carnapian realism about mathematical objects (outside set theory) I advocate in Part III unacceptable.

Third, I aim to provide a foundation for potentialist set theory which rests entirely on intuitively compelling principles which are subject-matter neutral and constrain the behavior of all objects (and thus perhaps accord with Frege's criterion for logicality[]. But I don't mean to claim that my foundational principles are analytic, cognitively trivial, or impossible for any rational being to doubt. I merely claim they're as prima facie obvious as the other axioms of set theory[10]. I also don't mean to suggest that the powerful proof transcendent facts about conditional logical possibility discussed in this book constitute some kind of metaphysical free lunch. I think complex considerations specific to set theory motivate going potentialist and inflating our fundamental ideology rather than our ontology in this case. But I don't claim any benefit with regard to Occam's Razor.

Fourth, we must distinguish my foundational project in this book, from a

---

is some hope that modal reasoning about potentialist set theory will let one prove *more*. So, in principle, the axioms of this book (or further equally 'clearly correct' seeming intuitive reasoning about my notion of conditional logical possibility) could motivate the addition of new axioms of mainstream set theory. I hope to explore this optimistic technical project elsewhere, but will say little more about it in this book.

[10] I take the axiom of choice to be prima facie obvious, despite the fact that can be doubted on grounds like the Banach-Tarski paradox. But readers who find choice less immediately appealing, can read this as a claim to justify Replacement from principles 'as prima facie obvious as the other axioms of *ZF* set theory' instead.

less ambitious sense in which one might try to provide a notion of set on which the axiom of replacement is "clearly true"[94]. Actualist philosophers have sometimes aimed to find a unified conception of set theory from which all the various ZFC axioms clearly follow – without worrying whether this conception itself is clearly coherent. This project can be valuable in various ways, e.g., in showing the naturalness and appeal of certain mathematical hypotheses (like proposed large cardinal axioms) which also follow from the relevant conception. However, finding such a unified conception doesn't suffice for my foundational project. For if the unifying conception isn't clearly consistent then, surely, it isn't clearly true (even on a view which allows mathematicians to introduce arbitrary coherent structures). So we haven't succeeded in justifying all theorems of ZFC set theory from premises that are clearly true.

Finally I should distinguish my foundational project from a *more* ambitious project involving justifying the ZFC axioms. Philosophers sometimes attempt to discover the most metaphysically joint carving successor to the naive concept of sets which generates Russell's paradox (something which might, e.g., be hoped to connect intimately with the true answer to the liar paradox). So, for example, you could ask whether the iterative hierarchy conception of sets is remotely on the right track, or whether the 'best' successor to naive set theory is something like Quine's New Foundations instead.

I don't think this more philosophically hopeful and speculative project is illegitimate. However, it's not relevant to my current project of providing a certain kind of foundation for the humbler mathematical knowledge we currently seem to have. For, if it turns out that, say, Quine's New Foundations is the 'true' set theory in the sense above (I take it) this wouldn't imply that typical late 20th century papers about set theory get things wrong — only that they don't talk about the most philosophically illuminating thing. So I'll try to show that potentialist translations of set theory provide a good Carnapian explication of set theoretic talk in typical 21st century mathematical contexts, and that (once translated) all theorems of ZFC set theory follow from principles that seem clearly true. But I won't take a position on claims about this more speculative project.

Chapter 2

Actualist Set Theory

In this chapter I will discuss the traditional, actualist, approach to set theory. I will review how the actualist faces problems articulating a categorical conception of the intended height of the hierarchy of sets (despite the existence of certain categoricity and quasi-categoricity theorems). I will then discuss how the actualist faces problems justifying the axiom of replacement from principles that seem clearly true.

2.1 Actualist Set Theory and The Iterative Hierarchy Conception

On a straightforward actualist approach to set theory, there are abstract objects called 'the sets', much as there are abstract objects called 'the natural numbers.' And we can ask: what sets exist? And what kind of structure do the sets have under the relation of membership?

Naively one might want to say that, for any formula $\phi(x)$, there is a set whose elements are exactly those objects that satisfy ϕ. But, as Bertrand Russell famously showed, this leads to paradox via the conclusion that there must be a set whose elements are exactly the sets which aren't members of themselves.

The (widely embraced) iterative hierarchy conception of the sets solves this problem by suggesting a different picture of what sets exist. On this picture we think about the sets as forming layers, with sets at a given layer in the hierarchy only being able to have elements which are available at previous layers. Each layer contains 'all possible sets' of elements given at prior layers and no two sets have exactly the same elements[1]. On this picture the height of a hierarchy refers to the 'number' of stages while the width refers to how many sets are introduced at each stage. One can spell out this idea of full width out by saying that

Definition 2.1.1 (Iterative Hierarchy - Full Width (IHW))**.** A full width iterative hierarchy (IHW) is a structure consisting of

- a well ordered series of levels (for reasons of tradition I will call these 'stages', but I don't mean to evoke any kind of temporal process) and
- a collection of sets 'available at' these stages, such that

[1] Note that there's been some discussion about whether extensionality follows from the iterative concept of sets or is something separate. But the worries I raise for actualists won't depend on the idea that our conception of the hierarchy of sets must be 'unified' in this strong sense. The question I will be pressing in sections 2.4 is merely whether we have a coherent conception of the hierarchy of sets (once the incoherence of our naive conception of the hierarchy of sets is recognized) that even seems to pick out a unique structure, not whether that conception is unified in the strong sense evoked above.

- At each stage there are sets corresponding to 'all possible ways of choosing' some sets available at lower levels (note that this can be stated straightforwardly in second order logic)

- Sets x and y are identical iff they have exactly the same members (extensionality).

One can think of the iterative hierarchy conception as specifying a structure for initial segments of the hierarchy of sets. If we adopt the idea of a hierarchy of sets, then the principles above specify an intended width for this structure. When we understand the demands above in terms of second-order logic, I'll refer to this idea as IHW_2 [2].

However, the principles above do not specify an intended height for the hierarchy of set[3].

Indeed, as we will now review there are important reasons for doubting that we have any coherent and adequate conception of absolute infinity (the supposed height of the hierarchy of sets). And, indeed, the version of potentialism I favor will wind up denying that there's, strictly speaking, a hierarchy of sets (hence anything for mathematical talk of 'the height of the hierarchy of sets' refers to[4]. But we will prove that speaking as if one were

[2] However, my preferred approach will reject the formalization in second order logic in favor of one $IHW_\Diamond$ using only the conditional logical possibility operator $\Diamond_{...}$ introduced in Chapter 4.I'll understand IHW loosely to be compatible both with a Boolos style two sorted conception as well as the standard cumulative hierarchy.

[3] We could add the principle that there is no last stage, as Boolos[13] does. But since there are many different logically possible well orderings which do not have a last element, e.g., ω, $\omega + \omega$ etc. this does still not give us a unique intended height.

[4] Instead we will analyze set theoretic talk as expressing potentialist claims about logical possibility, and extendability.

quantifying over sets and using first order logic is a harmless shorthand in ordinary mathematical contexts.

2.2 A Burali-Forti Problem

The problem, which I will present below, is not simply that it might be impossible to cash the notion of absolute infinity out in other terms. After all, every theory will have to take some notions as primitive. Rather, the problem is this.

- It's (prima facie) impossible for any collection of objects to satisfy a common naive conception of the intended height of the hierarchy of sets. Just as Russell's paradox shows that it's logically impossible for any collection of objects to satisfy the axioms of naive set theory. So our naive conception of absolute infinity (the height of the actualist hierarchy of sets) looks incoherent, not just not analysable.
- And, once we reject this naive conception, there's no obvious fall back conception that *even appears* to specify a unique height for the hierarchy of sets in a logically coherent way.

Specifically, a very common intuitive conception of the hierarchy of sets says that the hierarchy of sets goes 'all the way up' – so no restrictive ideas of where it stops are needed to understand its behavior. However, if the sets really do go 'all the way up' in this sense, then it would seem that they should satisfy the following naive height principle.

Naive Height Principle For any way some things are well-

> ordered by some relation R, there is an ordinal corresponding to it.

But, for example, the ordinals themselves are well-ordered, and there is no ordinal corresponding to this well-ordering, i.e., there is no ordinal which has the same order-type as the class of all ordinals. Thus (it would seem), the naive height ordering principle above can't be correct.

And it seems arbitrary to say that the hierarchy of sets just stops somewhere, if a suitable stopping point is not pinned down by something in our conception of the hierarchy of sets.

To clarify this worry, note that I'm not suggesting the actualist must think the hierarchy of sets 'must stop somewhere' in the sense that they must say there's a largest ordinal. There's no problem about saying that every ordinal has a successor (as indeed is required Boolos' version of the iterative hierarchy of sets). There's no problem about saying that for every set/ordinal x there's a strictly larger set/ordinal y. Nor is the issue that there could or should be 'sets beyond all the sets'.

Rather the problem is that the actualist takes there to be some plurality of objects (the sets) forming an iterative hierarchy structure (i.e. satisfying the description of the intended *width* of the hierarchy of sets above). And the following modal intuition seems appealing: for any plurality of objects satisfying the conception of an iterative hierarchy above (i.e., for any model of IHW), it would be *in some sense* e.g., conceptually, logically, combinatorially if not metaphysically possible for there to be a strictly lager model of

IHW which, in effect, adds a new stage above all he ordinals within the original structure together with a corresponding layer of classes. Note, I won't say more about how to spell this intuition out here, because each version of potentialist set theory discussed below (mine included) includes a way of sharpening the notion of possibility and extendabilty invoked here.

And it seems that the resulting structure generated would answer everything in our conception of the sets as well as the original structure did. For, once we've rejected the naive conception of the intended height of the hierarchy of sets above as inconsistent, we don't seem to have anything that even pretends to pick out a unique height.

Thus the actualist seems forced to say that the plurality of existing sets just happens to instantiate one possible/logically coherent structure satisfying the iterative hierarchy conception of sets rather than another conception which satisfies this conception equally well. They seem forced to say that the plurality of sets has some particular height, although nothing in their conception forbids it from being some even taller structure.

Saying that the hierarchy of sets just happens to stop at a certain point seems to violate intuitive principles of metaphysical parsimony. For one seems committed to positing an extra - otherwise entirely unmotivated - joint in reality, namely the height of the hierarchy of sets. One might also worry about the epistemology of this stopping point, and why we should think set theorists' reasoning about large cardinals etc. correctly reflects this brute fact about where the hierarchy of sets happens to stop.

The simplest response to this problem might be to find some other restrictive characterization of the sets (in particular, some other characterization of the intended height of the hierarchy of sets)[5]. However, there's no obvious fall back/replacement conception that even seems to pick out a unique structure. It's not clear that *any* precise intuitive conception of the intended height of the sets remains once the paradoxical well-ordering principle above is retracted. As Wright and Shapiro put it [108], all our reasons for thinking that sets exist in the first place appear to suggest that, for any given height which an actual mathematical structure could have, the sets should continue up past this height.

.

Moreover, the sets lose a substantial aspect of their appeal as a mathematical foundation if we can't capture all talk of coherent mathematical structures within set theory, i.e., via quantification over the sets or some set model that's at least isomorphic to the relevant mathematical structure. However, it is (at best) unclear whether we can do this if we accept actualism and say that the hierarchy of sets doesn't 'go all the way up' in the sense indicated above. Of course, by Gödel's completeness theorem for first-order logic, any consistent collection of first-order axioms will have a model. However our conceptions of mathematical structures (like, famously, the natural numbers) can include non-first-order notions, like second-order quantification. So the completeness theorem doesn't guaranteed that our conceptions of these structures will have 'intended' models in the hierarchy of sets (i.e.,

[5] Note that the axioms of ZFC and even ZFC_2 don't suffice to categorically determine the height of the hierarchy of sets.

models which treat their non-first-order vocabulary standardly).

One might further press this objection, by arguing as follows. If there were an actualist hierarchy of sets we could refer to, then we could also uniquely describe the possible structure which you would get by adding a single layer of classes to this hierarchy of sets. This structure is a legitimate topic for mathematical investigation, and yet this structure is not instantiated anywhere within the hierarchy of sets.[6]

Now, we could avoid the above worry about arbitrariness while securing a definite height for the hierarchy of sets, by simply *adding* some new idea about height to our current conception of the hierarchy of sets. For example, we could say that the sets are the shortest possible structure satisfying ZFC_2 (i.e., the sets up to the first inaccessible). This proposal is somewhat analogous to saying that the numbers are 'as short as can be' while being closed under successor and satisfying all the other first-order Peano Axioms – as we do when we take the natural numbers to satisfy induction.

However, making this kind of height-minimizing stipulation seems to fit badly with actual mathematicians' interest in large cardinals (which require the set theoretic hierarchy to extend far beyond the shortest model of ZFC). And, more generally, stipulating any height for the hierarchy of sets does nothing to help with the secondary worry above, that actualists shortchange the intended generality of set theory. Note that, if the actualist could plausibly claim to be able to define the height of the hierarchy of sets in terms of primitives they accept (possibly including the notion of absolute infin-

[6]See [56] Hellman for a version of this generality worry.

ity) sufficiently precisely to determine the truth-value for all set-theoretic statements then this worry wouldn't have any bite for them. However, even though people do use the term 'absolute infinity', this seems to be little more than a name for whatever height the hierarchy of sets has and not something they accept as a true meaningful primitive.

2.3 Categoricity and Quasicategoricity Arguments

2.3.1 McGee and Appeal to Ur-elements

With this worry about stating a precise conception of the hierarchy of sets (and avoiding arbitrariness) in place, let me quickly explain why two categoricity theorems which might seem to help the actualist don't help her.

In 'How we learn mathematical language'[80] Vann McGee advocates an interesting and influential conception of an iterative hierarchy of sets with ur-elements which might seem to help solve the problem of commitment to an arbitrary stopping point for the hierarchy of sets noted above.

However, I will argue that this is an illusion. Although McGee's characterization of a hierarchy of sets[80] solves the problem he is concerned with in that paper (addressing a certain kind of referential skepticism), it does not make the height of the actualist hierarchy of sets look any less arbitrary.

In [81] Van McGee defends realist claims that we can secure definite reference to the hierarchy of sets up to isomorphism (and thereby justify our presumption that all questions in the language of set theory have definite right answers) from a reference skeptical challenge.

Specifically he proposes an account of how creatures like us could count as having a definite conception of the sets up to isomorphism, given the presumption that we can secure definite realist reference for other kinds of vocabulary, and (it will be important to note) that we are somehow able to quantify over everything (sets included).

McGee explains how we can secure (the effect of) definite reference to second-order quantification and thus uniquely describe the intended width of the hierarchy of sets, via a story about schemata which I won't summarize here. Then he suggests that we can pin down the intended height of the hierarchy of sets by considering a conception of a hierarchy of sets *with ur-elements.*

The idea of set theory with ur-elements is simply to allow sets to have elements that aren't sets. A common way of developing set theory with ur-elements is to keep the core idea of an iterative hierarchy of sets described above (with each layer containing 'all possible subsets' from the lower layers), but take the lowest level of the hierarchy of sets to include sets corresponding to all ways of choosing from among all the objects that aren't sets (e.g., elephants, billiard balls, electrons, marriages and the like), rather just the empty sets. Note that the hierarchy of sets with ur-elements includes all pure sets. Thus, uniquely pinning down a hierarchy of sets with ur-elements would suffice to pin down a hierarchy of pure sets as well.

The following Ur-element Set Axiom follows from the statement above. It says that there's a set which contains, as elements, all the objects that aren't sets.

Ur-element Set Axiom (U) $(\exists x)(Set(x) \wedge (\forall y)(\neg Set(y) \rightarrow y \in x))$

McGee shows that we can (in a sense) pin down the intended height of this hierarchy of sets with ur-elements, if we accept the axiom above.

Specifically, McGee proves that $ZFC_2 + U$ (the result of adding the above ur-element principle to second-order ZFC set theory) has a property which he calls 'quasi-categoricity'[7]. Given any single choice of a total domain (what you are quantifying when you quantify over everything *including the sets*) there cannot be two non-isomorphic (with respect to $\in$) interpretations of set theory which both: choose 'sets' from within this domain, take quantifiers to range over this whole domain and make McGee's $ZFC_2 + U$ come out true (while interpreting all logical vocabulary standardly).

McGee's theorem ensures that we couldn't have a single universe containing both a hierarchy of red sets and a hierarchy of blue sets, such that both hierarchies satisfy the constraints imposed by $ZFC_2 + U$ on their relationship to the total universe (red sets and blue sets included). So it does the job McGee wants: answering skeptical challenges about definite reference to the hierarchy of sets (up to isomorphism), on behalf of a Platonist who presumes that there is an actualist hierarchy of sets and grants that we can somehow unproblematically quantify over everything (sets included).

However, this theorem does nothing to address the objection to actualism raised at the beginning of this chapter: that actualists seem committed to

[7] One might worry about the above axiom on the basis of Uzquiano's [117] proof that McGee's axioms for set theory with urelements are incompatible with certain axioms of mereiology, but I leave this question aside as the concerns I will be raising are unrelated.

an additional and arbitrary joint in reality – a point where the hierarchy of sets just happens to stop.

For McGee's theorem does not imply that we have any beliefs which logically necessitate (and thereby make non-arbitrary) facts about where the hierarchy of sets happens to stop. As McGee himself points out, the conception of sets he articulates is **not** categorical; the beliefs about the sets which he invokes are compatible with many different possibilities about how large the total universe of sets is.

Indeed, it's crucial to notice, McGee's theorem doesn't show that ZFC+U is 'quasi-categorical' on the following stronger (and, to my mind, more natural) sense. It doesn't show that, fixing the facts about what non-set objects there are, any hierarchy of sets satsifying ZFC+U must have a certain unique structure. For this is almost certainly[8] false. You can take one possible scenario containing a hierarchy of sets satisfying $ZFC_2 + U$ within a total universe of a certain size, add some sets to the top of this hierarchy, and therefore to the universe, (without changing any facts about the non-sets) and get another possible scenario satisfying $ZFC_2 + U$.

Thus McGee's theorem doesn't pin down a unique intended structure for the hierarchy of sets, or abolish arbitrariness by explaining why the hierarchy of sets stops at some particular point. It just shows that you couldn't have two non-isomorphic hierarchies of sets satisfying the above conception within the same universe.

[8] Here I speak assuming the possibility of inaccessible cardinals going arbitrarily high up as most do.

One could use McGee's conception of sets with ur-elements in a slightly different way which *would* block the arbitrariness worries for actualism I've pressed above, as follows. Assume that our use of non-mathematical vocabulary to pins down the intended interpretation of certain non-mathematical kind terms. We could specify the intended height of the hierarchy of sets by saying that (in effect) the hierarchy of sets stops *as soon as it can* while satisfying $ZFC_2 + U$.

Unfortunately, however, this proposal faces the same worries about making the hierarchy of sets too small, which arose for the idea that we could just pick a restrictive conception of the sets in section 2.2. It also suggests the height of the hierarchy of sets might be contingent and that the result of physical and metaphysical investigation into how many non-mathematical objects there are should have bearing on facts about pure set theory in a way that seems potentially uncomfortable.

2.3.2 Martin

Similarly Martin's categoricity theorem about set theory in [79] might at first sound like it helps the actualist with her arbitrariness/lack of a definite conception worry, but actually does not.

Indeed, as it happens, Martin seems to positively endorse a version of this worry. For in [78] he distinguishes five ingredients in our conception of the hierarchy of sets as follows.

> The modern, iterative concept has four important components:

> 1. the concept of the natural numbers;
>
> 2. the concept of sets of xs;
>
> 3. the concept of transfinite iteration;
>
> 4. the concept of absolute infinity.
>
> Perhaps we should include the concept of Extensionality as Component (0).

And then he expresses the following reservations about whether we have a definite coherent notion of absolute infinity.

> so I am using the term “absolute infinity” for the concept that is the fourth component of the concept of set. One can argue that the concept is categorical, and that any two instantiations of the concept of set (of the concept of an absolutely infinite iteration of the sets of x’s operation) have to be isomorphic. But it is hard to see how there could be a full informal axiomatization of the concept of set. There are also worries about the coherence of the concept. People worry, e.g., that if the universe of sets can be regarded as a “completed ” totality, then the cumulative set hierarchy should go even further. Such worries are one of the reasons for the currently popular doubts that it is possible to quantify over absolutely everything. I am also dubious about the notion of absolute infinity, but this does not make me question quantification over everything.

In [79] Martin argues against plenitudinous anti-objectivist 'multiverse' approaches to set theory (like [51]) on which certain set theoretic claims Φ are not determinately true or false for the following reason.

> Multiverse Idea: The platonic realm of mathematical objects includes many different (non-isomorphic) hierarchies of sets. There's no unique intended V, even up to width. Rather for each hierarchy V in the multiverse is expanded by some larger one which adds, e.g., a 'missing' subset of the natural numbers V (So, we might note, none of these hierarchies can answer our conception IHW of the width of the hierarchy of sets above). Some of these Vs make Φ true and others make Φ false. And all of them are (absent specific mathematical choice to 'work in' a particular hierarchy of sets) equally intended.

Martin argues against this multiverse proposal (and, I think, poses a powerful challenge to it) as follows. He notes that if we accept a certain conception of the hierarchy of sets (and the principles below), it follows that there could not be two different non-isomorphic hierarchies of sets 'the red sets' and 'the blue sets' could not simultaneously exist.

- a 'uniqueness' principle: all sets are extensional. That is, if there are two distinct sets x and y (even in two different putative hierarchies!), then there must be some object which is an element of x but not y or vice versa. Thus, for example, there can be only one set $\{Mars, Venus\}$.

- a conception of the hierarchy of sets, including (among other more familiar elements) the following height closure principle: if a set exists, then any hierarchy of sets containing the elements of that set must contain the set itself

Martin shows that it follows from the principles above, essentially by induction, that there can't be two different hierarchies of sets. Any two putative hierarchies of sets satisfying the conditions above, must agree on their urelements, and then on their first layer and their second layer etc.

One can call this a categoricity result. But note that it doesn't imply that it's logically or metaphysically necessary that any collections of objects which satisfy our conception of sets must have a certain (unique) structure. Rather, it merely shows that there can't be two distinct *actual* set-theoretic hierarchies. For example, Martin's argument doesn't rule out the possibility that there could be some description of an ordinal ϕ_κ, such that it would be logically possible to have a structure satisfying our conception of the sets containing an ordinal satisfying ϕ_κ but also logically possible to have such a structure which didn't contain any ordinal satisfying ϕ_κ. It merely shows that we couldn't have two actual hierarchies of sets, one of which contains ϕ_κ while the other does not.

2.4 A Problem Justifying Replacement

In addition to the worry above (about whether we have a coherent conception of the intended height of the hierarchy of sets), set theoretic actualists also face a problem about justifying the axiom schema of replacement. They must make it plausible that whatever unique height (and hence structure) they think hierarchy of sets has, satisfies replacement.

Informally, the axiom schema of replacement tells us that the image of any set under a definable (with parameters) function is also a set. More formally, let ϕ be any formula in the language of first-order set theory whose free variables are among $x, y, I, w_1, \ldots, w_n$. We can think of the formula $\phi(x, y)$ (and choice of parameters) as specifying a definable function (with parameters) taking x to the unique y such that $\phi(x, y)$. Then the instance of axiom schema of replacement for this formula ϕ says the following:

$$\begin{aligned} &\forall w_1 \forall w_2 \ldots \forall w_n \forall a \\ \text{(Repl)} \qquad &\qquad \forall x(x \in a \to (\exists! y)\phi(x, y, w_1, ...w_n)) \\ &\qquad \leftrightarrow \exists b \forall x\big(x \in a \to \exists y(y \in b \land \phi(x, y, w_1, ...w_n))\big)\big] \end{aligned}$$

So replacement says: whenever some first-order formula defines a function on a set a, i.e., associates each element x of a with a unique y, there is a set b equal to the image of a under this function. In other words, the hierarchy of sets extends far enough up that all the elements in the image of a can be collected together.

As Boolos points out in [14], the axiom of replacement imposes a kind of closure condition on the height of the hierarchy of sets which doesn't obviously follow from the iterative hierarchy conception of the sets above, even if we add the claim that there is no last stage. For consider $V_{\omega+\omega}$. This structure satisfies both of the above assumptions; each layer of sets contains sets corresponding to all ways of choosing sets lower than it, and there isn't a last layer. However, it doesn't satisfy Replacement, since you could take the set ω (formed at layer $V_{\omega+1}$) and write down a function ϕ which associates 1 the $\omega+1$, 2 with $\omega+2$ etc. Then for each x in ω there's a y in $V_{\omega+\omega}$ satisfying $\phi(x,y)$. But there isn't any set b in $V_{\omega+\omega}$ which collects together the image of every member of ω. That set b is only formed at a $V_{\omega+\omega+1}$. This raises a worry about how to justify Replacement, and (indeed) whether mathematicians are justified in using it at all.

So (even if we take for granted that there are objects satisfying the iterative hierarchy conception of sets), if we want to justify use of the ZFC axioms, a question remains about how to justify the axiom of replacement.

There has been much interest and sympathy with this worry in the subsequent literature. For example [94] Putnam writes, "Quite frankly, I see no intuitive basis at all for . . . the axiom of replacement. Better put, I do not see that a notion of set on which that axiom is clearly true has ever been explained."

And, more recently, in a discussion of the history of set theory Michael Potter remarks that, "it is striking, given how powerful an extension of the theory replacement represents, how thin the justifications for its introduction were" [9], and then reports that "In the case of replacement there is, it is true, no

[9] He supports this assessment by quoting "Skolem... gives as his reason that 'Zermelo's axiom system is not sufficient to provide a complete foundation for the usual theory of sets', because the set $\{\omega, P(\omega), P(P(\omega)), ...\}$ cannot be proved to exist in that system; yet

widespread concern that it might be, like Basic Law V, inconsistent, but it is not at all uncommon to find expressed, if not by mathematicians themselves then by mathematically trained philosophers, the view that, insofar as it can be regarded as an axiom of infinity, it does indeed, as von Neumann ... said, 'go a bit too far'"[90].

To my knowledge, four main (actualist) strategies for justifying replacement are currently popular.

First, people have tried to justify using the axiom of replacement 'extrinsically' in much the way that one would justify a scientific hypothesis, by appeal to its fruitful consequences, arguing it helps prove many things we independently have reason to believe and hasn't yet been used to derive contradiction or consequences we think are wrong[10].

The first strategy only provides extrinsic justification, whereas it's at least prima facie appealing to expect central principles of set theory which are used without comment to have intrinsic justification, and this expectation seems common in other areas of mathematics. For example, it seems that

this is a good argument only if we have independent reason to think that this set does exist according to 'the usual theory', and Skolem gives no such reason. Von Neumann's ... justification for accepting replacement is only that, ' in view of the confusion surrounding the notion 'not too big' as it is ordinarily used, on the one hand, and the extraordinary power of this axiom on the other, I believe that I was not too crassly arbitrary in introducing it, especially since it enlarges rather than restricts the domain of set theory and nevertheless can hardly become a source of antinomies."'.

[10]See [64] which quotes Gödel's description, "Even disregarding the intrinsic necessity of some new axiom, and even in case it had no intrinsic necessity at all, a decision about its truth is possible also in another way, namely, inductively by studying its "success", that is, its fruitfulness in consequences and in particular in "verifiable" consequences, i.e., consequences demonstrable without the new axiom, whose proofs by means of the new axiom, however, are considerably simpler and easier to discover, and make it possible to condense into one proof many different proofs."

everything we want to say about the natural numbers follows from (say) our second-order conception of the natural numbers. Now if it turns out that adequate intrinsic justification cannot be given, it might be reasonable to accept extrinsic justification (for we do this in the sciences, after all). And probably we will reach a point with, e.g., large cardinal axioms where extrinsic justification is all we can provide.

However, one might hope to do better with regard to the ZFC axioms, which are treated as quite secure and used to provide a foundation/explication of normal mathematical claims that we are very confident in. And, even if these strategies succeed in providing *some* justification for using the axiom of replacement, I think it will be agreed that none of them account for the kind of intrinsic convincingness we usually expect (and hope for) from mathematical axioms.

Second, people have tried to justify replacement by appeal to a kind of inference to the best explanation along the following lines. Russell's paradox tells us that not all pluralities of objects can form a set (there isn't a set of all sets that aren't members of themselves). So if there are any sets, there should be a principled division between those pluralities of objects can form a sets and those which can't. But sets don't have that many features. So (one might think) size is the only natural choice for the limitation on what pluralities count as sets and it should be the *only* such limitation [90]. As Michael Potter puts it, we should accept the following Size Principle, "If there are just as many Fs as Gs, then the Fs form a collection if and only if the Gs do." (which implies Replacmenet) because

> "[A] collection is barely composed of its members: no further structure is imposed on them than they have already. So... what else could there be to determine whether some objects form a collection than how many there are of them? What else could even be relevant?"

As regards the second strategy (inference to the best explanation, along the lines Potter proposes), I will suggest that sets do have *some* other features than their size which could be used to give an explanation for why there isn't a Russell set analogous to the one Potter gives.

In particular note that, given the iterative hierarchy conception of sets (which Potter accepts) each set will have the property of first being generated at some ordinal level α. This feature of sets as a fairly natural and principled one. One can think of it as reflecting how many layers of indirect and metaphysically derivative object existence (given the common idea that sets are in some sense metaphysically dependent on their elements, not vice versa [11]) one has to go through to arrive at that set.

So, rather than hypothesizing (with Potter) that the iterative hierarchy of sets stops at a certain point because ascending any further would require collecting objects which are *too plentiful* to form a set, couldn't we just as well hypothesize that the iterative hierarchy of sets stops somewhere because any further sets formed would have to occur *too high up* in an

[11] See, for example [12] for the a development of the intuition that the existence of Socrates' singleton is to be grounded in the existence of Socrates and depends on that, in a way that the existence of Socrates does not depend on the existence of his singleton, and use of this intuition to motivation a notion of grounding which is distinct from metaphysically necessary covariation and supervenience.

iterative hierarchy (i.e., one would have to ascend through too many layers of abstraction/metaphysical dependence to form a set from the relevant elements).

To the same (rather fanciful) extent that we can imagine that the rubber band holding together the elements of a sets just happens to be too small to collect any plurality of elements of a certain size κ, we could imagine that the power of lower level sets to ground the existence of higher level sets and thereby indirectly to ground the existence of still higher level sets etc. eventually becomes too attenuated to allow any further sets to be formed at some height α.

So how do we know there's an upper bound to the *sizes* sets can have vs. an upper bound to the *rank* they can have?

One might also object to Potter's methodology more generally, on the grounds that even philosophers who are happy to use this kind of metaphysical inference to the best explanation suggested by Potter's justification don't usually take applying this method to justify the great confidence and certainty we feel we give to typical mathematical results.

Third, people have proposed a kind of justification for replacement by noting it follows from a set theoretic reflection principle[12]. I take this proposal (and the one that follows) to typically arise from the attempt to find a unified conception of the sets from which the ZFC axioms follow (whether or not that conception is obviously true or coherent) as per §1.6, rather than any

[12] My summary of this approach follows [64]

attempt to derive the axiom of replacement from something that seems more obviously true. But I will discuss both proposals for completeness.

Informally, the idea behind reflection principles is that the height of the universe is "absolutely infinite" and hence cannot be "characterized from below". A specific reflection principle will assert that any statement ϕ in some language that's true in the full hierarchy of sets V is also in some smaller V_α. This ensures that one cannot define V as the unique collection which satisfies ϕ (or the shortest such collection) since there will be a proper initial segment V_α of V that satisfies ϕ.

More formally, once accepts first order reflection/second order reflection etc. insofar as one accepts all instances of the following schema, where ϕ is a first order/second order etc. formula.

Reflection Schema For any objects $a_1, \ldots, a_n$ in V_α, we have $\phi(a_1, \ldots, a_n) \leftrightarrow V_\alpha \models \phi(a_1, \ldots, a_n)$.

If one accepts first order reflection then one can justify Replacement[13].

This third strategy (justification by appeal to a reflection principle) is *somewhat* attractive. For, as Koellner reviews in[64] one can motivate reflection principles[14] by Gödel's idea that the total hierarchy of sets (V) should be impossible to define. For reflection principles (in effect) say that anything that's true of the whole hierarchy of sets will also be true in some proper initial segment of it. If some instance of a reflection principle failed (so there

[13] [fix and ref]

[14] Different reflection principles correspond to different classes of sentences being reflected. For instance, you might think only first-order sentences reflect or first-order formulas with parameters or second-order sentences etc.

was some fact about the whole hierarchy of sets that didn't reflect down to be true of a proper initial segments of the sets) then we could (in a sense) define the hierarchy of sets by saying it is the shortest[15] iterative hierarchy structure satisfying this claim. 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 taken in [a] more and more generalized and idealized sense."[16]

I admit that the idea in the quote above has a kind of elegance and provides a kind of internal justification for reflection (as opposed to the external justification by consequences evoked above).

However it's not obvious (or not as obvious as we'd naively hope foundational axioms for mathematics could be) that there could be a structure satisfying the intuition behind reflection (or even second order reflection) together with our other expectations about the hierarchy of sets (e.g., the other ZFC axioms, and the width conditions discussed above).

Also, to the extent that Gödel's idea in the quote above motivates the first-order Reflection principle used to justify Replacement above, it would seem to also motivate third order reflection, some instances of which (as Koellner notes in the article cited above) have been shown to be inconsistent[101]. So one might think that justifying Replacement by merely noting that it follows from reflection doesn't provide enough justification[17].

[15] That is, the sets satisfy the non-reflected claim but no initial segment does.

[16] This is quoted from [118] in [64].

[17] One might also hope that one could use such reflection schemas to solve the first

Fourth, philosophers like Boolos [14, 16] justify replacement from a size principle. (Speaking informally) the idea is to say that some plurality of objects forms a set if and only if it is 'small' where the latter means that its members can't be bijected with the total universe. This principle justifies replacement, because the set you get by applying replacement to a set u must be the same size as u or smaller. .

But, just as with Reflection, it's not as clear as one would like that it would be coherent for there to be a structure with the intended width of the hierarchy of sets that satisfies this property together with the axiom of infinity[18].

problem above. For example, one could say that the hierarchy of sets as the *shortest* structure satisfying second order reflection. However the above mentioned worries above about why we should think it's logically coherent for there to be a structure satisfying the relevant form of Reflection remain. We also have a limitative proposal, generality of set theory intuitions evoked above. And we would have to find a version of Reflection weak enough to avoid the incoherence of third order Reflection but strong enough that set theorists would be willing to constrain themselves to studying the shortest hierarchy of sets satisfying it. Note that appeal to reflection principles (as at least plausible enough to be interesting to study) *is* popular with some set theorists. But this is not to say that set theorists accept such principles as intrinsically obvious or as a definition of the intended height of the hierarchy of sets in the limitative sense proposed above, as opposed to merely an appealing hypothesis about the height of the hierarchy of sets.

[18] One might even argue that if one had sufficient a priori confidence in Boolos' 'size of the total universe principle', this would provide reason to reject actualism. For, one might think, our conception of the intended height of the hierarchy of sets should be something which we are capable of grasping without appeal to how many non-mathematical objects happen to exist (or be useful to talk in terms of). But suppose we have some such intrinsic grasp of the intended structure of the pure hierarchy sets which pinned down that it stopped at some point α, how could we be rationally be sure that there wouldn't turn out to be (or it wouldn't turn out to be useful to think in terms of) strictly more non-set objects? If there were we'd have a violation of the principle above, for the sets would not be bijectable with the total universe, so there would have to be a set of all sets leading to contradiction. A potentialist like the one I develop has an easier time reconciling the idea that we can grasp facts about set theory directly, without appeal to any facts about what non-set objects there happen to be, and yet the potentailist translation of 'there is a set of all physical objects' (i.e., it would be logically possible for there to be a standard width iterative hierarchy with ur-elements structure containing objects that behave like a set and a bijection from the Fs) comes out knowable a priori) Given the modal logical analog to the axiom of choice i.e. Choice, all objects are isomorphically map-able to some ordinal.

So, to summarize the discussion of different actualist strategies for justifying replacement above, we get the following picture. In order to justify the level of confidence we have in set theory, and particularly Replacement, (as well as for aesthetic reasons) we would like our set theoretic axioms to follow from some simple intuitive conception which strikes us as prima facie clearly logically coherent.

For instance, we think of number theory as describing the sequence built by starting at 0 and continuing to add successors 'as long as is needed to ensure that there is no last natural number, but no longer' in a sense which can be cashed out via the second-order axiom of induction. And we can think of the reals as describing a line extending to infinity in both directions without gaps (i.e., such that it's impossible to add any further 'number' anywhere on the line without it being equal to a real[19]). In both these cases our conception of the mathematical structure seems to flow from a single unified conception that's intuitively consistent.

The iterative hierarchy idea sketched in §2.1 plausibly specifies the width of the hierarchy of sets in a way that's logically coherent (on its own). But just assuming that the sets satisfy this width requirement (or even that adding that there's no last stage to the hierarchy of sets) doesn't suffice to justify replacement. Adding principles like Reflection or Boolos' size principle to our conception would ensure that our conception of the intended structure of the sets implies replacement (and hence perhaps that if there are sets then

[19]One can think of a Dedekind cut which doesn't correspond to a real number as a kind of gap, i.e., a vertical line passing through the x-axis that somehow misses every real number.

they satisfy replacement). However we have little or no reason to think this enlarged concepion is cohernet. So it provides little justification for thinking that the axiom of replacement is even consistent with the other principles about the hierarchy of sets (hence little justification for thinking it's true).

In the next few chapters I will argue that moving to a potentialist approach to set theory lets us do better with regard to both the arbitrariness and justification problems above.

2.5 Indefinite Extensibility

But, before I go on to the development and defense of potentialism, let me end this chapter by quickly saying something about the limits of the argument above.

Many other philosophers interested in potentialism about set theory have also explored more general versions of potentialism, which go further and reject the idea that we have a definite conception of the structure of the natural numbers or the width of the hierarchy of sets. Thus one might wonder if there is a principled reason for taking a potentialistic approach to the height of the hierarchy of sets but not to the width of the hierarchy of sets or the natural numbers.

In the remainder of this chapter I will answer the above question. I'll explain why I think the above motivation for height potentialism about set theory doesn't generalize in the ways just mentioned. And I will contrast the claims I've made about our *lacking a coherent categorical* conception of an actualist hierarchy of sets above with Dummett's famous – and famously obscure – remarks about indefinite extensibility.

2.5.1 Height Potentialism And No More

In a nutshell, I think my limited potentialism is motivated in a principled way by the fact that our naive conception of the height of the hierarchy of sets gives rise to a Burali-Forti paradox, while no similar paradox seems to arise from taking the appearance that we have a coherent conception of things like the natural numbers, the width of the hierarchy of sets or full second-order quantification at face value.

Here's another way of thinking about the disanalogy. One can fairly concretely imagine an ordinal-like-object above any well ordered plurality of ordinals, and a layer of set-like-objects above any plurality of sets. We can specify exactly how $\leq$ and $\in$ would relate the new sets/ordinals to all the old sets/ordinals previously considered.

And it's prima facie plausible that the structure we imagine forming by extending any given plurality of ordinals has as good a claim to containing all the objects that satisfy our conception of 'the ordinals'/'the sets' as the original structure. For our conception of the ordinals/sets doesn't seem to include any (coherent) negative conditions which say that the height of the hierarchy must stop at a certain point.

But we can't do the same thing with our concepts of 'full' second-order quantification (aka arbitrary subsets of a given collection), natural number and real number. Perhaps, in a sense, it's intuitive that, for any collection of natural numbers (finite or infinite) we can imagine a strictly larger *vaguely* number-like object. For we can always imagine adding (something like) a successor or a limit ordinal after all numbers within any collection of numbers. However, our grasp of the natural numbers does very centrally include such a principle saying the numbers must stop at a certain point, namely the second-order induction axiom! We think the numbers are (so to speak) as *few as can be*[20] while containing 0 and the successor of everything they include, and that for this reason any property which applies to 0 and applies to the successor of everything it applies to must apply to all the numbers. The same goes for the concept of full second-order quantification/all possible subsets of a given collection. We have no positive intuition about how to generate, for any given collection of sets of cats a new set-of-cats like object which is distinct from all the ones previously considered.

2.5.2 Contrast with Dummett

It may be helpful at this point to contrast my arbitrariness problem for actualism with Michael Dummett's influential arguments about indefinite extensibility in *The Seas of Language*. In [30], Dummett raises something very much like the Burali-Forti worry I pressed above concerning the height of the hierarchy of sets. For example he expresses something that sounds somewhat like the arbitrariness worry I pressed against the actualist above.

> If it was... all right to ask, "How many numbers are there?", in

[20] Here I mean 'few' in an order type sense not a cardinal sense. Maybe it would be better to say that the natural number structure is as short/small as can be while satisfying this condition

> the sense in which "number" meant 'finite cardinal', how can it be wrong to ask the same question when "number" means 'finite or transfinite cardinal'? A mere prohibition leaves the matter a mystery. It gives no help to say that there are some totalities so large that no number can be assigned to them. We can gain some grasp on the idea of a totality too big to be counted, even at the stage when we think that, if it cannot be counted, it does not have a number; but, once we have accepted that totalities too big to be counted may yet have numbers, the idea of one too big even to have a number conveys nothing at all. And merely to say, "If you persist in talking about the number of all cardinal numbers, you will run into contradiction", is to wield the big stick, not to offer an explanation.[21]

And one might say that both of us reject standard actualist set theory on the grounds that our conception of sets is, in some sense, 'indefinitely extensible'. However, Dummett is concerned with indefinite extensibility in a different sense than I am. Specifically, I reject standard (actualist) platonsim about set theory because our concept of sets and ordinals is 'indefinite extensibility' in the following strong sense:

> **Strong Indefinite Extensibility** We have a positive intuition that for any hierarchy of sets/ordinals *there could be* there could be a strictly larger one which matches our iterative hierarchy conception of sets/ordinals equally well.

[21] [30] pg. 439

In contrast, Dummett rejects standard Platonist set theory because our concept of sets is 'indefinite extensible' in this weaker sense:

> **Weak Indefinite Extensibility** For any collection of them *we can definitely imagine* (which he says he will start by presuming means any *finite* collection) this collection can be extended so as to contain extra things which would also fall under our conception of that structure[22].

To support this reading, note that Dummett argues that the concepts of natural numbers are 'indefinitely extensible' by (seemingly) assuming that all totalities of numbers we can form a definite conception of collect numbers from 0 to n for some n. His story about how to extend an arbitrary totality of natural numbers (that we can definitely conceive of) is simply the following.

> given any initial segment of the natural numbers, **from 0 to n**, the number of terms of that segment is again a natural number, but one larger than any term of the segment.

Similarly the argument Dummett takes to show that our concept 'real number' is indefinitely extensible is simply Cantor's diagonal argument that any countable plurality of real numbers must be leaving some real numbers out.

Indeed Dummett explicitly notes that he's making these assumptions (of finiteness and countability) in the quote below and (unsurprisingly) recognizes they will strike opponents as question begging.

[22]Dummett writes, "[A]n indefinitely extensible concept is one such that, **if we can form a definite conception of** a totality all of whose members fall under the concept, we can, by reference to that totality, characterize a larger totality all of whose members fall under it."[30] pg 440].

> A natural response is to claim that the question has been begged. In classing real number as an indefinitely extensible concept, we have assumed that any totality of which we can have a definite conception is at most denumerable; in classing natural number as one, we have assumed that such a totality will be finite. Burden-of-proof controversies are always difficult to resolve; but, in this instance, it is surely clear that it is the other side that has begged the question.

Dummett goes on to defend this burden of proof claim, by arguing that it's mysterious how a definite conception of an infinite structure could be communicated and the burden of showing such communication is possible falls on his opponent. Much can and has been said about whether this succeeds, and how to understand Dummett's infamously "dark"[] notion of indefinite extensability[74, **?**]. However all that matters for my purposes is that Dummett's arguments from the weak indefinite extensability of the natural numbers and real numbers don't even pretend to show the strong indefinite extensability of these notions. They don't pretend to show that, for *any* totality of objects related by some relation R in the way we believe the natural numbers to be related by successor, it would be logically coherent to have a strictly larger structure that accords with our conception of the natural numbers equally well.

Thus Dummett's reason for worrying about the sets arguably applies to the natural numbers and real numbers (any *finite* collection of these will be missing a number which could be added) etc. while (we've just seen above that) mine doesn't.

Philosophically speaking, I suspect these different 'indefinite extensibility' worries arise from different philosophical projects and background assumptions as follows.

I take both the naive intuition that we mean something definite by both 'all possible subsets' and 'all the way up' at face value, until Burali-Forti paradox shows the latter is contradictory. Since no analogous paradox seems to arise for 'all possible subsets' I'm happy to invoke this notion in expressing a conception of the natural numbers etc.

In contrast, Dummett starts from a more skeptical/cautious position and asks to be shown how one could 'convey' a definite concept of structures to someone who starts out only understanding finite collections. And he prima facie doubts that you could do so by, e.g., giving an operation like adding one and talking about closing under it or (as I would prefer) or appealing to a modal notion of 'all possible' subsets which applies to infinite collections.

In this way I take the pages above to express a principled reason for doubting that we have a consistent categorical conception of the hierarchy of sets which (unlike Dummett's concerns) doesn't similarly call into doubt the appearance that we have a definite categorical conception of the natural numbers, the real numbers, or of what it takes for there to be a layer of classes including classes witnessing 'all possible ways of choosing' from some plurality of independently understood objects like cats or sets within some iterative hierarchy [23].

[23] Now, (tangentially to the main argument above) Dummett would presumably chal-

lenge me to say how we could grasp the notion of 'all possible subsets' of a given infinite collection which we naively seem to grasp. And, in a nutshell, I'd answer that we can latch onto a notion of logical possibility which (we will see below) suffices to categorically describe the numbers and sets in the same way (whatever it is) that we can latch on to a notion of objective physical possibility/law. For example it might be that we get both notions by making certain core good inferences (e.g., the actual to possible 8.2.1 and uniform relabling 8.5.1 principles below the case of logical possibility, and some other kind of extrapolation in the case of physical possibility) which in a way under-determine which modal notion we mean and then benefiting from reference magnetism. Thus I suspect that Dummett's worry either (despite protests to the contrary) comes down to an argument from some principle of manifestability which would call reference to realist physical possibility/law facts into doubt as well reduces to mine or. However I won't pursue this argument here, because my present aim is only to explain how my worry differed from Dummett's, not to answer his worry.

Chapter 3

Putnamian Potentialism: Putnam and Hellman

Let us now turn to potentialism, a different approach to set theory which promises to answer (at least) the arbitrary height worry raised for the actualist above. There are two broad schools of potentialism which I will, following [73], call Puntamian and Parsonian potentialism.

Parsonian potentialists take there to be objects called sets, just like the actualist. And they take set theory to describe how a literal hierarchy of sets could grow. So they think that the (pure) sets exist contingently, in some sense.

In contrast, Putnamian potentialists don't take set theory to describe actual or possible hierarchies of special objects called sets. Instead they understand set theory to make claims about how it's possible for some objects to (under some relations) satisfy *set theoretic axioms*, like ZFC_2 or IHW_2, and how it would be possible for such structures to be extended, i.e., how those objects (or objects with the same set-theoretic structure) could be supplemented by further objects to form a longer iterative hierarchy. for these objects, or some with the same structure under the relevant relations, to exist alongside other objects jointly forming a longer iterative hierarchy structure.

In this chapter I will review existing work in the Putnamian school, which I favor. Later, in chapter 5 I will discuss rival Parsonain proposals and compare them to my proposed formulation.

How does this work? As noted above, a Putnamian will paraphrase set theoretic sentences as making claims about how there could be (objects that, when considered under certain relations, jointly have the structure of certain) standard width initial segments V_α of the total hierarchy of sets V, and how such structures could be extended, i.e., exist as a standard width-initial segment of a longer V_β. They systematically replace quantification over the sets with modal claims about how such structures can be extended as follows.

A set theorist's sentence of the form $(\exists x)(\phi(x))$, where ϕ is quantifier free (e.g., $(\exists x)(x = x)$), gets formalized as saying that it would be possible for there to be (objects with the structure of) a standard width initial segment of the hierarchy of sets containing an object x satisfying ϕ (in this case $x = x$).

Correspondingly, a set theorists' $(\forall x)(\phi(x))$ claims, where ϕ is quantifier free, says that it's necessary that[1] any object x within a standard width initial segment of the hierarchy of sets has the property ϕ.

What about set theorists' sentences with more quantifiers, like a claim of the form $(\forall x)(\exists y)\phi(x, y)$, where ϕ has no quantifiers?

The Putnamian takes this to say that it's necessary that for any (objects forming a) standard width initial segment V and object x within it, it's possible to have larger initial segment V' extending V, containing an object y, such that $\phi(x, y)$. And the same pattern continues for more logically complex sentences.

But, unsurprisingly, there has been much philosophical disagreement and debate about how to fill the details of this picture out. For example, what is the correct notion of possibility to employ here? What does it take for some things to form a standard width initial segment? And what logical tools should we use to articulate our answer to the above?

In this chapter I'll review Putnam and Hellman's development of Putnamian potentialist set theory and highlight some unresolved issues.

3.1 Putnam

In [95] Hilary Putnam sketches a way of thinking about set theory in terms of modal logic: as talk about what 'models' of set theory are, in some sense, possible and how such models can be extended.

[1] That is, it's necessary that if some objects have the structure of a standard with initial segment then any x in that structure satisfies ϕ.

He introduces a notion of being a standard model of set theory, which is a model of set theory closed under subsets, i.e., a hierarchy of sets having full width and no infinite descending chains under $\in$[2]. Putnam says that we can 'make this notion concrete' by thinking of models as physical graphs consisting of pencil points (or the analog of pencil points in space of some higher cardinality) and arrows connecting these pencil points. And he "ask[s] the reader to accept it on faith" that we can express the claim that some model is standard in this way "using no 'non-nominalistic' notions except the '$\Box$"' (where $\Box$ denotes the logical necessity operator).

With this notion of a concrete model in place, Putnam suggests that we can understand set theoretic statements as claims about what such models are possible, and how they can be expanded. For example, he proposes that we can paraphrase a set theoretic statement of the form '$(\forall x)(\exists y)(\forall z)\phi(x, y, z)$' where ϕ is quantifier free, as saying that, if G is a standard concrete model, and p is a point within G, then it is possible that there is a model G' which extends G, and a point y within G' such that necessarily, for any model G'' which extends G' and contains a point z, $\phi(x, y, z)$ holds within the concrete model G''. And we can treat arbitrary quantified statements in set theory in an analogous fashion.

Putnam then suggests that adopting this potentialist approach to set theory can help us dispel the kind of arbitrariness and indefinite extendability

[2]Specifically, Putnam writes "[A concrete] model will be called standard if (1) there are no infinite-descending 'arrow' paths; and (2) it is not possible to extend the model by adding more "sets" without adding to the number of "ranks" in the model. (A 'rank' consists of all the sets of a given-possibly transfinite-type. 'Ranks' are cumulative types; i.e., every set of a given rank is also a set of every higher rank. It is a theorem of set theory that every set belongs to some rank.)"

worries I discussed in section 2.2 above. For, adopting this approach lets us understand set theoretic talk without imposing or positing arbitrary limits on the size of structures (as we would do if we just stipulated a point at which the hierarchy of sets stopped, or inferring that it must stop somewhere) in a way that seems faithful to our intuitions about the generality of set theoretic reasoning[3]. As Putnam puts it,

> "[W]e have a strong intuitive conviction that whenever As are possible, so is a structure that we might call 'the family of all sets of As.' ...from the standpoint of the modal-logic picture ... the Russell paradox ... shows that no concrete structure can be a standard model for the naive conception of the totality of all sets; for any concrete structure has a possible extension that contains more "sets." (If we identify sets with the points that represent them in the various possible concrete structures, we might say: it is not possible for all possible sets to exist in any one world!) Yet set theory does not become impossible. Rather, set theory becomes the study of what must hold in, e.g., any standard model for Zermelo set theory."

I think Putnam is right that his proposal indicates an appealing style of response to the worries about arbitrary stopping points for the hierarchy of sets indicated above. And (as we will see) it has inspired many other philosophers. However, this proposal is (explicitly) sketchy on certain formal and philosophical points. For instance, Putnam doesn't provide any criteria

[3]In particular, (before thinking about the paradoxes) we'd hoped for set theory to be general in the sense that every possible structure will have a copy somewhere in the sets.

for what it would take for some collection of arrows and pencil points to form a standard model and he asks the reader to "accept it on faith that the statement that a certain graph G is a standard model for Zermelo set theory can be expressed using no 'non-nominalistic' notions except the '□'."[4]

And, philosophically, Putnam says very little about the notion of 'mathematical possibility' which he intends to capture with the □, and seems to vacillate between a purely mathematical understanding of necessity and a physical understanding.

For example, he writes (brackets in original), "assuming that the notions of mathematical possibility and necessity are clear [and there is no paradox associated with the notion of necessity as long as we take the '□' as a statement connective (in the degenerate sense of "unary connective") and not...as a predicate of sentences], I wish to employ these notions to try to give a clear sense to talk about 'all sets'." [?] However, at earlier points Putnam talks like conceivable constraints on how many physical objects like pencil points and lines could fit into physical space are relevant, and makes assumptions about this which philosophers like Parsons[87] and Tait[114] have been unwilling to grant, e.g., Putnam says, "I assume that there is nothing inconceivable about the idea of a physical space of arbitrarily high cardinality; so models of this kind need not necessarily be denumerable, and may even be standard."

Additionally, Putnam advocates potentialism as merely one possible and

[4]Here nominalistic notions are ones that aren't committed to the literal existence of mathematical objects.

helpful 'perspective' on mathematics and claims that it is, in some sense, equivalent to a more familiar actualist understanding of set theory, which only appears to be incompatible with it. But cashing this idea out clearly requires serious and disputable metaphysics[5].

Furthermore, it's not clear that saying both perspectives are equally good is compatible with honoring Putnam's potentialism-motivating intuition that "whenever As are possible, so is a structure that we might call 'the family of all sets of As"[]. We seem forced to *either* say that the idea that for any structure there could be a larger one is only true 'from the potentialist perspective' on mathematics or to say that it is true simpliciter, even from the actualist perspective.

The former position can feel a little mysterious and unsatisfying, but the latter is uncomfortable for two reasons. First (like more straightforward forms of actualism) it involves positing arbitrariness in mathematical reality by saying the actualist hierarchy of sets just happens to stop somewhere, though it could go on further. Second, it's not clear (even at a very loose intuitive level) how talking about any such actualist hierarchy could be equivalent to a practice of modal set theory which considers arbitrary logically possible extendability[6].

---

[5] See, for example, John Burgess' vigorous objections to Putnam's stance in [17].

[6] Perhaps one could say that the actualist hierarchy is the smallest standard width structure whose truth conditions for all first-order logical claims agree with those provided by the potentialist set theory.

3.2 Hellman

In [56, 46] and [?] Hellman develops Putnam's ideas about potentialist set theory as part of a larger purely nominalist philosophy of mathematics, and solves or avoids many of the problems above.

First, Hellman drops Putnam's suggestion that actualist and potentialist approaches to set theory are (somehow) supposed to be two equally good perspectives on the same thing and instead merely advocates and develops the potentialist approach. I will follow suit.

Second, Hellman provides a somewhat clearer picture of what the key modal notion $\Diamond$ in (his version of) Putnam's potentialist set theory is supposed to mean, saying that it's supposed to express a primitive modal notion of logical possibility. However he does relatively little to describe this notion. He does say that, "[when evaluating logical possibility] we are not automatically constrained to hold material or natural laws fixed." So it may be logically possible that $(\exists x)(\text{pig}(x) \wedge \text{flies}(x))$, but physically impossible. And he adds that, "we are free to entertain the possibility of additional objects — even material objects — of a given type.". So, for example it's logically possible that there are infinitely many objects even if there are actually only finitely many objects. So, for example, it's logically possible for there to be say 2^{2^ω} cats, even if it's not metaphysically possible for there to be so many cats. This (arguably) lets us avoid concerns about limitations on the cardinality of space unduly limiting the range of possible models considered above. Beyond this remark, however, Hellman just suggests that his applications of logical possibility will make the notion he has in mind clear. I will echo this move as well (while saying much more about the relevant notion of logical possibility and why we should accept it as a modal primitive below).

Hellman also does a lot to fill in the other promissory notes left by Putnam's sketch. He cashes out Putnam's appeal to 'standard models' of set theory by saying that standard models are models which satisfy ZFC_2 (i.e., the version of standard ZFC set theory which replaces the inference schemas of replacement and comprehension with corresponding second-order axioms)[7]

Note that where Putnam spoke of models of "Zermelo set theory" (which doesn't include replacement) Hellman talks abut hierarchies satisfying second order ZFC, which do satisfy replacement. That is, Hellman takes the initial segments whose possible extensions potentialism considers to *themselves* satisfy ZFC_2. ,

Now *if* one accepts the relevant large cardinal axioms, then there's a sense in which this change makes no difference. For (it turns out [8]) that potentialist translations taking initial segments to themselves satisfy ZFC_2 will be logically equivalent to translations involving initial segments that satisfy much weaker requirements IHW_2 or Zermlo set theory. But this is a poor fit to my current project of justifying the potentialist version of ZFC from intuitively compelling principles[9] For note that to infer even the simplest existential

[7] So, for example, ZFC expresses comprehension via an axiom schema which contains an axiom for every formula ϕ in the language of set theory. In contrast, by using second-order logic one can state a single comprehension axiom as follows $(\forall x)(\forall C)(\exists y)(\forall z)(z \in y \leftrightarrow z \in x \wedge C(z))$. The same goes for the first-order axiom schema of replacement and its second-order analog.

[8] See **??**

[9] I gather Hellman [**?**] chooses to go this way in an attempt to bring out a kind of

claim in set theory (e.g., to say that there is a set that is self-identical), we would need to know that it was logically possible for a structure to satisfy ZFC_2. And the logical coherence of a hierarchy of sets satisfying second order ZFC is by no means obvious, especially in the context of our current doubts about replacement.

One *might* also feel that requiring the initial segments being extended to satisfy ZFC_2 or even constitute a 'standard model of Zermelo set theory' (rather than merely satisfying our conception of being an intended width hierarchy IHW above, e.g., IHW_2) is slightly unnatural[10]. In chapter 2 I tried to paint the following picture. We seem to have a precise and consistent conception of the intended width of the hierarchy of sets, but (as we see when deriving contradiction from the Naive conception of absolute infinity in §2.2) no such conception of its intended height. Now one might say: the point of potentialism as a solution to the arbitrariness problem, is to solve this problem of heights. So potentialist set theory should talk about how iterative hierarchies of standard width could be extended, rather than imposing any height constraints. But I admit that perhaps this is a matter of taste.

Finally, Hellman also makes one more change in [55] which I want to highlight because my own proposal will wind up being closer to Putnam's original

analogy between replacement and large cardinal axioms, something which I don't attempt here.

[10]Note, I say 'unmotivated', not illegitimate. As discussed in §3.3, I take it mathematicians are free to study something like potentialist set theory (how any structures satisfying certain axioms can be extended by structures studying other axioms), just as they are free to study objects satisfying any logically coherent pure mathematical axioms in a non-potentialist spirit.

proposal in this regard. When Putnam talks about the modal perspective on mathematics, he considers possibility of objects being related by **specific first order relations** as per certain set theoretic axioms. So, for example, we might consider the possibility that the pencil points form an intended model of Zermelo set theory when considered under the relation arrows(,):'an arrow points from.. to ...' . If we followed Hellman in requiring our hierarchies to satisfy ZFC_2, this would amount to saying that all axioms of ZFC_2 become true when you replace 'set' with 'point' and 'element of' with 'arrows'. However, he notes that any relations of the right arity will do [11]. We could translate a given sentence of set theory equally well by talking about how it would be logically (or logico-mathematically in whatever sense Putnam has in mind) possible for the pencil points to arrow one another or the angels to admire one another.

In contrast, Hellman interprets set theoretic claims purely in terms of second-order quantification in [55]. That is, instead of saying something about how

[11] For example on pages 10-11 of [92] he writes "Let 'AX' abbreviate the conjunction of the axioms of the finitely axiomatizable subtheory of first-order arithmetic just alluded to. Then Fermat's last theorem is false just in case '$AX \supset \neg$ Fermat' is valid, i.e., just in case

(1) $\Box(AX \supset \neg$ Fermat)

Since the truth of (1), in case (1) *is* true, does not depend upon the meaning of the arithmetical primitives, let us suppose these to be replaced by "dummy letters" (predicate letters). To fix our ideas, imagine that the primitives in terms of which AX and ¬Fermat are written are the two three-term relations "x is the sum of y and z" and "x is the product of y and z" (exponentiation is known to be first-order-definable from these, and so, of course, are zero and successor). Let $AX(S,T)$ and ¬Fermat(S,T) be like AX and ¬Fermat except for containing the "dummy" triadic predicate letters S, T , where AX and ¬Fermat contain the constant predicates "x is the sum of y and z" and "x is the product of y and z." Then (1) is essentially a truth of pure modal logic (if it is true), since the constant predicates occur "inessentially"; and this can be brought out by replacing (1) by the abstract schema: (2) $\Box[AX(S,T) \supset \neg\text{FERMAT}(S,T)]$ -and this is a schema of pure first-order modal logic."

it's logically possible for penciled points to arrow one another, we talk about the possibility of there being second-order classes and functions objects X and f such that ZFC_2[set/X,∈/f].

Putting this all together, in [55] [12], Hellman defines the claim that some second-order collection X and relation[13] f quantifiers, (X, f) form a natural model of set theory as follows.

$$(X, f) \text{ form a natural model of set theory iff } ZFC_2^X(^{\epsilon}_{f})$$

where the expression on the right hand side of the biconditional is what you get by starting with ZFC_2 and then uniformly replacing all occurrences of ϵ with f, and reinterpreting all quantifiers as ranging over the objects in X rather than the sets. He then paraphrases singly quantified set theoretic statements (i.e., those of the form $\exists x\phi(x)$ where ϕ is quantifier free) as

$$\Diamond(\exists X)(\exists f)[(X, f) \text{ form a natural model of set theory } \wedge (\exists x)\phi(x)^X(^{\epsilon}_{f})]$$

For readability he then uses quantification over variables of the form V_i to abbreviate quantification over X_i, f_i which form a natural model of set theory, with claims of the form $z \in V_i$ standing for the claim that $z \in X_i$, for the relevant X_i. So, for example, the paraphrase for $\exists x(x = x)$ gets written as

[12] I suppress one detail of Hellman's paraphrase strategy (his separate treatment of set theoretic statements involving only restricted quantification) which makes no difference to the philosophical arguments being made here. See [56] chapter 2 section 2.

[13] Here I follow Hellman in using f as a second order **relation** rather than function quantifier as its form might suggest

$$\Diamond(\exists V)(\exists x)(x \in V \wedge x = x)$$

And he goes on to paraphrase claims with a single universal quantifier and multiple nested quantifiers as per the pattern indicated in the beginning of this chapter. So $(\forall x)(x = x)$ will get translated as

$$\Box(\forall V)(\forall x)(x \in V \rightarrow x = x)$$

And let us say that a model of set theory $V_2 = (X_2, f_2)$ extends another model $V_1 = (X_1, f_1)$ (written $V_2 \geq V_1$) iff X_1 is a subclass of X_2 and f_1 is the restriction of f_2 to X_1.This amounts to requiring that (X_1, f_1) and (X_2, f_2) pick out structures which relate to each other like initial segments of an actualist hierarchy V_α and V_β where $\alpha \leq \beta$.

Then Hellman circa [55] would translate the set thoeretic sentence $(\forall x)(\exists y)(x \in y)$ as follows.

$\Box(\forall V_1)(\forall x)[x \in V_1 \rightarrow \Diamond(\exists V_2)(\exists y)(y \in V_2 \wedge V_2 \geq V_1, \wedge x \in y)]$

In later work [46] Hellman modified this view slightly, as motivated by his nominalism and famous Quineian sentiment that second-order logic is ontologically committal [**?**] (so accepting second order comprehension commits one to abstract objects). He notes that quantifying over all pluralities xx automatically lets you simulate second order X quantification, via the strategy indicated in the appendix of David Lewis' *Parts of Classes*[70]. Accordingly, he proposes to rewrite all the paraphrases above plurality of objects xx (in effect) satisfying second order set theory, written ZFC_2^{xx}, rather that, rather than second order objects X, f doing so[14].

I have some doubts about the success of this move. In particular, I'm not convinced that this use of mereology to simulate second order quantification can be combined with taking the $\Diamond$ to express logical possibility rather than metaphysical possibility. For reasons that will become clear below[15], I don't think that the axioms of mereology are logically necessary. If logical possibility ignores metaphysically necessary constraints on how many concrete objects can exist in space and time, shouldn't it ignore the metaphysically necessary laws of mereology too? Thus, I think that employing this strategy to formulate potentialist set theory (rather than just modally paraphrasing talk of smaller structures like the numbers and the reals as Hellman suggests in [55]) would reawaken the problems about the metaphysical possibility of arbitrarily large cardinalities of objects noted above. And, as we will see, my approach also eliminates use of second order quantification in favor of a notion of logical possibility with, arguably, a stronger claim to ontological innocence than plural quantification.

But I won't dwell on this issue more here[16] as I'm not a nominalist myself.

[14] In even later work [**?**] Hellman switches to a two sorted view (more like Parsonian views discussed below) above where we have stages s, and plural quantification over pluralities of stages ss, and also plural quantification over objects xx which will play the role of sets in satisfying ZFC_2

On this view these stages is understood in terms of there actually being a growing sequence of objects in a Parsonaian sense. But the claim that some xx satisfy $ZFC_2^{xx,ss}$ (will now include the claim that these sets same heights as some stages) will still be understood structurally as saying that certain axioms are satisfied.

[15] If we take logical possibility to be interdefinable with validity in the way I'll advocate in §3.5, the logical contingency of mereology seems to follow.

[16] I do consider switching to my potentialist framework as a friendly suggestion to advocates of nominalism (as I suggest in [6])

3.3 Attractions

All together, I think adopting potentialist set theory in the manner developed by Putnam and Hellman has significant appeal. As noted above, it helps solve the arbitrariness problem which actualists face regarding the height of the set-theoretic hierarchy.

Additionally, unlike Parsonain approaches, adopting this view doesn't require us to make sense of the idea of pure mathematical objects like sets exist contingently, or explain what contingent features of reality (if any) determine the height that the set-theoretic hierarchy actually/currently has.

However, merely adopting Hellman's Putnamian potentialism doesn't suffice to secure our foundational aim of justifying set-theoretic theorems from obvious seeming assumptions.

Hellman does proves a version of the main theorem one needs, to vindicate standard first order reasoning about set theory.

$ZFC \vdash \phi$ then $\phi^{\Diamond}$

However, the premises he uses in this proof aren't (and aren't claimed to) seem clearly true. For instance, in [55] Hellman simply assumes that the translation of replacement into a potentialist context as an axiom and explicitly flags that it is not intuitively obvious.

Hellman also provides a justification for the use of the ZFC axioms from the point of view of an actualist who accepts certain assumptions about higher set theory. Hellman's justification goes like this. Assume that actualist set theory is true and there are cofinally many inaccessible cardinals. On this assumption, we can re-interpret (Hellman's preferred version of) potentialist claims as claims about what initial segments of the true hierarchy of sets exist. Then it is a theorem that, for each first-order set theory sentence ϕ, this re-interpretation of the potentialist translation of ϕ will be true iff the original sentence ϕ is true. Thus, since ZFC_2 is presumably true of the actualist hierarchy of sets, the potentialist translation of these claims will also come out true.

But (as Hellman himself notes), the justification he provides for the ZFC axioms is not satisfactory from a potentialist point of view, because it requires that we assume the existence of an actualist hierarchy of sets. Additionally, we must also assume that this acualist hierarchy has co-finally many inaccessible cardinals satisfies (a somewhat controversial large cardinal axiom). So, even from an actualist point of view, you might say that Hellman's argument defending the use ZFC justifies the more obvious on the basis of the less obvious. So, while Hellman's justification might be a useful rhetorical tool for convincing actualists, it doesn't provide a justification for using the ZFC axioms which the potentialist can accept.

In later work, Hellman experiments with other justifications for replacement. But it should be noted that in doing this his aim is only to motivate unifying principles by showing key set-theoretic beliefs follow from a single natural hypothesis (as per **??**), not significantly justify replacement itself. So the hypotheses from which Hellman derives replacement don't seem any more clearly true than the potentialist translation of replacement and often much less so. For example, in [] Hellman considers a modal reflection principle, which would justify potentialist replacement but, just as in the actualist case, seems no more obvious than replacement itself[17] And in [?], Roberts argues this principle is inconsistent with other axioms Hellman should plausibly endorse.

3.4 Problems for Current Putnamians

With this picture of the current state of Putnamain potentialist set theory in mind, I'll now discuss two potential problems about Hellman's system.

The first worry concerns whether we have a clear enough grasp of Hellman's notion of logical possibility ($\Diamond$). I think this worry can be answered and will attempt to provide such an answer below. The second worry concerns controversies about the meaning and truth-value of basic claims in the language of quantified modal logic. I'll use this worry to motivate a switch to my preferred version of Putnamian potentialism' in chapter 4.

[17]He motivates this principle by considering the following statement of a potentialist replacement principle, "The mathematical possibilities of ever larger structures are so vast as to be "indescribable": whatever condition we attempt to lay down to characterize that vastness fails in the following sense: if indeed it is accurate regarding the possibilities of mathematical structures, it is also accurate regarding a mere segment of them, where such a segment can be taken as the domain of a single Structure." However he notes this is inconsistent, and tries restricts its application to things consistent with ZFC_2. But this principle doesn't seem any more obvious that the reflection principles invoked by actualists discussed in §??.

3.5 Acceptably Clear Modal Notion?

On Logical Possibility

So, let me begin by clarifying and motivating the notion of logical possibility that Hellman appeals to (and I will also appeal to).

We seem to have an intuitive notion of logical possibility which applies to claims like $(\exists x)(\text{red}(x) \wedge \text{round}(x))$ and makes sentences like the following come out true.

- It is logically possible that $(\exists x)(\text{red}(x) \wedge \text{round}(x))$
- It is not logically possible that $(\exists x)(\text{red}(x) \wedge \neg\text{red}(x))$
- It is logically necessary that $(\forall x)(\text{red}(x)) \rightarrow \neg(\exists x)(\neg\text{red}(x))$.

This notion of logical possibility is interdefinable with validity. An argument is valid if and only if it's logically impossible for all its premises to be true and its conclusion to be false. And it is (roughly) what's analyzed by saying some theory has a set theoretic model[18] (modulo concerns about size, as noted in the appendix below). It concerns whether some state of affairs is allowed by the most general 'subject matter neutral' laws of how there can be some pattern of objects standing in relations of various arities (in something like Frege's sense of logical laws being subject matter neutral[]).

Philosophers representing a range of different views of mathematics have made use of this notion and are comfortable applying it to non-first order

[18] When considering non-first order sentences we might specify that this model must treat all logical vocabulary standardly, so that, e.g., Henkin models of second order quantification are not allowed

sentences.

To evaluate whether a claim ϕ is logically possible (in this sense), we hold fixed the operation of logical vocabulary (like $\exists, \wedge, \vee, \neg$), but abstract away from any further metaphysically necessary constraints on the application of particular relations. Thus, we consider all possible ways for relations to apply (including those ways that aren't definable). For example, it is logically possible that $(\exists x)(\text{Raven}(x) \wedge \text{Vegetable}(x))$, even if it would be metaphysically impossible for anything to be both a raven and a vegetable.

We also abstract away from constraints on the size of the universe, so that $\Diamond(\exists x)(\exists y)(\neg x = y)$ would be true even if the actual universe contained only a single object.

Not Reducible to Set Theory

Because (as noted above) the notion of logical possibility is interdefinable with validity, I think nearly all my readers will accept that claims about logical possibility are meaningful.

However, at first glance, one might argue that claims about logical possibility are merely shorthand for claims about the existence of set theoretic models. And if one identified logical possibility the notion of logical possibility with claims asserting the existence of set theoretic models, then we'd have (at least) an uncomfortable regress, and one couldn't use the notion of logical possibility in formulating potentialist set theory to solve the arbitrariness problem above.

Luckily however, there are strong independent reasons pointed out in [49, 53, 15] (see also [39] 2.3 and Etchemendy [34]) for not doing this. Many philosophers have argued, as follows, that we shouldn't identify claims about logical possibility with claims about set theoretic models.

The claim that what's actual is logically possible is central to the above notion of logical possibility (interderivable with validity), if anything is[19]. However if we think about logical possibility in terms of set theoretic models, then the actual world is strictly larger than the domain of any set theoretic model (e.g., because it contains all the sets), so it's not prima facie clear why we should assume that what can't be satisfied in any set theoretic model isn't actually true. Thus we seem to antecedently grip a notion of logical possibility (interdefinable with validity) on which it's an open question whether every logically possible state of affairs has a set theoretic model.

Now it is *currently* possible for mathematicians talking about *first order logical sentences* to replace talk of logical possibility with talk of set theoretic models via the completeness theorem for first order logic[20] However, as Boolos puts it, "it is rather strange that appeal must apparently be made to one or another non-trivial result in order to establish what ought to be obvious: viz., that a sentence is true if it is valid"[15].

I also feel (with Boolos) that, "one really should not lose the sense that

[19] For an argument to be valid surely at least requries that it doesn't actually lead from truth to falsehood.

[20] The completeness theorem shows that all syntactically consistent first order theories have models [**?**]. And the notion of logical possibility is intuitively 'sandwiched between' syntactic consistency and having a model (anything that has a model must be logically possible, and anything that's logically possible must be syntactically consistent), so this shows that all three notions apply to exactly the same first order logical sentences [40].

it is somewhat peculiar that if G is a logical truth, then the statement that G is a logical truth does not count as a logical truth, but only as a set-theoretical truth.", and so reject cashing out claims about failures of logical truth/validity in terms of logically contingent claims about the existence of certain objects (even mathematical objects). To foreshadow slightly, following Boolos' suggestion, I will treat the $\Diamond$ of logical possibility a primitive modal operator, and furthermore *logical* operator whose meaning must be held fixed when we're evaluating claims about logical possibility and entailment. Thus we will affirm that facts about logical possibility are themselves logically necessary truths.

Contrast With Other Modal Notions

It may be useful to note how the above notion of logical possibility differs from three vaguely similar modal notions in the literature, namely Tarskian re-interpretability, metaphysical possibility and conceptual possibility.

The notion of logical possibility is (potentially) less demanding than the notion of truth under some Tarskian reinterpretation, for approximately the reason discussed above (and emphasized in [35]). Certain scenarios might be genuinely logically possible but require the existence of more objects than actually exist, and hence not permit any Tarskian reinterpretation. For, Tarskian reinterpretations of a sentence must still take the sentence's quantifiers to range over some collection of objects in the actual world.

The notion of logical possibility is also prima facie less demanding than the notion of metaphysical possibility[21]. For, as Frege noted, the laws of logic hold at all possible worlds. Yet it would seem that statements like $(\exists x)(\text{Raven}(x) \wedge \text{Vegetable}(x))$ can require something which is logically possible but metaphysically impossible.

Finally, the notion of logical possibility is also less demanding than the notions of idealized conceivability and conceptual possibility at issue in debates over philosophical zombies and in Chalmers' *Constructing the World*[20] (and are, inconveniently, sometimes also labeled logical possibility). For the notion of conceptual possibility reflects something like ideal a priori acceptability. So, when evaluating whether it is conceptually possible that ϕ we have to preserve all analytic truths associated with relations occurring in ϕ. In contrast (as I have noted above) logical possibility abstracts away from all such specific features of relations. Thus, for example, if we assume it is analytic that $(\forall x)(bachelor(x) \rightarrow male(x))$, then it will be logically possible but *not* conceptually possible that $(\exists x)(bachelor(x) \wedge \neg male(x))$.

In view of all the points above, I take it that there's no problem in (and indeed significant independent motivation for) accepting that we have primitive modal notion of logical possibility which will do the work the Putnamian potenailist wants it to regarding avoiding worries about metaphysically necessary limitations on size.

Talk of arguments' validity (in some sense) seems to be widely understood

[21] I want to leave it open the possibility that on some kind of ideal logical analysis, logical possibility turns out to be the same thing as metaphysical possibility. I'm just noting that we have a concept of logical possibility independent of this assumption, and that this suffices to give an attractive account of set theory.

and useful. And, by the arguments above, cashing validity claims out by appeal to a primitive modal notion of logical possibility, rather than attempting to reduce it to a notion of having a set theoretic model or truth under some Tarskian reinterpretation, seems like the wisest course.

3.6 Quantified Modal Logic

Now I want to raise a second problem which, I'll argue, does make it difficult to combine use of Hellman's paraphrase strategy the main project of this book: justifying set theory via principles that are as intuitively obvious seeming as possible.

This problem concerns the infamous controversialness of quantified modal logic. Putnam and Hellman both quantify in to the $\Diamond$ of logical possibility (or whatever other modality is used to cash out potentalist set theory). That is, they use sentences like $\exists x \Diamond R(x)$, where the logical possibility operator is applied to a formula with free variables. But, there are significant controversies about the truth value (and/or meaning) of even very simple sentences involving quantifying in to the $\Diamond$ of logical or metaphysical possibility.

Quinean Qualms

Most radically, Quine famously argued against quantifying in to modal contexts all together. I take Quine's main problem with quantifying in, in [96], to be that he dislikes the "Aristotelian essentialism" of saying that some properties belong to an object like the number 7 essentially (e.g., being less than 9) while others apply only contingently (e.g., being the number of planets). After all, taking there to be such an abundance of facts about essences can seem like positing a bunch of arbitrary and unneeded metaphysical facts. But perhaps these concerns are less severe if we specify that we're only talking about logical possibility, because objects' logical essences will be (somehow) 'minimal'.

Contingent Objects

More influentially at the moment, there's debate among philosophers who accept quantified modal logic (and quantifying in) about whether everything exists necessarily. In most (reasonably strong) quantified modal logics we can prove the following claim which seems to say that everything exists necessarily. In particular, if we take $\phi(x)$ to be $x = x \rightarrow (\exists y)(y = x)$ we easily see that $(\forall x)\phi(x)$ is logically true and thus infer $(\forall x)\Box\phi(x)$, i.e., $(\forall x)\Box[x = x \rightarrow (\exists y)(y = x)]$. We can thus infer the following – a sentence that seems to say everything exists necessarily.

$$(\forall x)\Box(\exists y)(y = x)$$

Hellman follows Kripke [55][66], by saying that familiar principles from sentential modal logic like the necessitation rule and K in S5[22] only apply to complete sentences in quantified modal logic. And perhaps this is intuitively motivated. We wouldn't want to say it's logically necessary or a tautology that $x = x$, because formulas with free variables aren't even sentences and thus lack truth values.

[22] These rules are, respectively: if $\vdash A$ then $\vdash \Box A$) $(\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

However, this response is controversial. For example, an alternative approach would be to allow quantifying in, but use free logic[83][23].

Also note that on Kripke's[67] approach sentences like $(\exists x)\Diamond[\neg \mathrm{Fox}(x) \wedge (\forall y)\mathrm{Fox}(y)]$ are true if there are any contingent objects (a conclusion which can't be easily avoided[24]), a consequence which Williamson[120] points out is fairly counterintuitive[25].

Necessary Distinctness

Next, disagreement can arise about whether all pairs of things that are actually distinct are necessarily distinct. For example in [42] Fine considers making this assumption and whether it can address Quinean worries about (logical) essences mentioned above.

> There are, of course, familiar Quinean difficulties in making sense of first-order quantification into modal contexts when the modality is logical. Let me here just dogmatically assume that these

[23]Switching to an free logic would let us block the above argument by blocking the initial proof that '$(\exists y)(y = x)$', rather than the application of necessitation to this formula in the last sentence as free logics neither assume that all singular terms refer to members of the domain nor that the domain is non-empty.

I think this strategy is prima facie quite appealing, because it would allow us to capture the intuitive logical possibility of entirely empty domains. However because, as a matter of sociological fact, no free logic is currently widely accepted (and because avoiding quantifying in makes the implications of adding any given axiom more obvious), I have preferred to sacrifice intuitions about empty domains and use classical first order logic rather than arguing for new views on both first order logic and set theory in this book.

[24]For note that, when considering the truth value of $\mathrm{Fox}(x)$ under an assignment of 'x' to some contingent object o that doesn't exist at some possible world w, it seems we must say that x isn't in the extension of 'Fox' at w, (since it would be weird to insist that objects that don't exist at w were nonetheless foxes) and hence that $\neg Fox(x)$ should be true under this assignment.

[25]While this debate is commonly conducted in terms of metaphysical possibility, it naturally raises similar concerns for logical possibility.

> difficulties may be overcome by allowing the logical modalities to 'recognize' when two objects are or are not the same.Thus
>
> $\Box\forall x\Box(x = y \rightarrow \Box x = y)a$
>
> and
>
> $\Box\forall x\Box(x \neq y \rightarrow \Box x \neq y)$
>
> will both be true though, given that the modalities are logical, it will be assumed that they are blind to any features of the objects besides their being the same or distinct.

But (to the extent that we have any grip on quantifying in to logical possibility) this assumption is dispuatable. For example some have argued that it's metaphysically (and hence presumably logically possible) for there to be two people who could have been one person. Suppose that two people are formed by a contingent event of person splitting e.g., a Star Trek transporter malfunction or a brain getting split in half and each side regrowing. One might think these people are distinct but could have been identical[26][27].

[26] I take this point from [?]

[27] Readers may recall that Fine makes some further points against using the notion of logical possibility for his purposes in that paper: charitably and non-paradoxiccally formulates debates about 'absolute generality' i.e. whether it is possible to quantify over everything in a certain strong philosophical sense, and develop arguments that, for any sets you are quantifying over, Russell's paradox shows that you could be interpreted as quantifying over more things. He argues that you can't because one can't see how any domain could be inextensible in the sense of logical possibility. But, of course, Hellman would accept the latter claim. And accepting it poses no problem for Hellman's use of logical possibility to provide a Putnamian potentialist reconstruction of set theory, which has no committment to there being an interesting sense in which there could be more sets but not more dogs.

Metaphysical Shyness?

Additionally in [73], Linnebo formulates a worry specific to Hellman, concerning the possibility of a kind of metaphysical or logical shyness. He writes, "Do we really know that there cannot be 'metaphysically shy' objects, which can live comfortably in universes of small infinite cardinalities, but which would rather go out of existence than to cohabit with a larger infinite number of objects?" This existence of such 'shy' objects would pose a problem for Hellman, because it could block us from saying that every plurality of objects forming a hierarchy of a certain kind could be extended in a certain way.

Linnebo also notes that if Hellman's notion of logical possibility allows for an analog to metaphysically incompatible objects (e.g., two metaphysically possible knives formed by joining a single handle with different blades) this can make certain assumptions Hellman uses to justify the existence of potentialist translations ZFC come out false.

Paraphrasing sentences of set theory with modal sentences that quantifying in to the $\Diamond$ of logical possibility forces us to consider when objects from one logically possible world are identical to or counterparts of objects in another. We are forced to ask whether, for some particular object, *that very object* could count as persisting in a world where the total universe has some cardinality, or some other possible object exists.

What to Do?

These controversies can raise doubts about whether our intuitions about quantifying in are reliable[28] and whether we can choose axioms for modal logic which are powerful enough to justify potentialist formalizations of the ZFC axioms but clearly and (fairly) uncontroversially true, in the way we'd like foundational mathematical axioms to be.

One could debate about whether the disagreements above are best understood as a philosophical disagreement about a proposition (e.g., that everything exists necessarily) or as showing that we don't have a good grip on what quantifying in means or that the formalism of quantified modal logic (that allows quantifying in) means different things to different people. But for my purposes either option would be a sufficient reason to avoid formulating our foundational modal axioms (used to justify set theory) in terms of quantifying in.

In general one might try to solve this kind of problem by stipulating that sentences which quantify in to the $\Diamond$ should be understood as having whatever meaning is necessary to make certain axioms true. But note that, for the purpose of formalizing set theory (as evenly modestly truth value realistically construed), this is approach won't do. For insofar as we need there to be proof transcendent facts about set theory, we can't just say that any interpretation of our $\Diamond$ quantified modal statements that satisfies certain

[28] My proposed account of set theory is compatible with taking Williamson to show that any modal notion *which allows quantifying in* (such as metaphysical possibility) must have a fixed domain – provided one thinks it doesn't make sense to quantify in to logical possibility. Of course, it's not compatible with taking Williamson to show that every modal notion must have a fixed domain.

axioms (sufficient to justify potentailist translations of set theoretic claims) is equally intended. We have to try to latch on to an intuitively meaningful notion, about which truth can outrun proof.

Instead, I propose to solve the above problem in a different way: by eliminating quantification in to the $\Diamond$ of logical possibility, as we will see in Chapter 4.

Chapter 4

Overview of My Proposal

Let me now turn to my preferred form of Putnamian potentialism, which will let us avoid the above (§3.6) problem of controversies about quantified modal logic.

In this chapter, I'll introduce my key conditional logical possibility operator, (which generalizes the independently motivated logical possibility operator discussed in §3.5), and discuss how using it to formulate potentalist set theory[1] is helpful. However, I'll delay actually using this notion to paraphrase set-theoretic claims until chapter 14.

In §4.1 I'll briefly motivate my basic approach to reformulating potentialist set theory. In §4.2.1 I'll introduce and clarify the key notion of notion of conditional logical possibility. In §4.3 I'll clarify the main advantages I take reformulating Putnamian potentialist set theory in terms of conditional

[1] I first advocated doing this as a way to remove redundancies from Hellman's modal structuralism in [8]

logical possibility to have. Then in §4.4 I'll clarify how the paraphrases I ultimately propose will differ from Putnam's and Hellman's, and note a further advantage.

I am advocating one choice of primitive, while another is much more familiar. I ask that readers evaluate this choice of primitives on the basis of philosophical fruitfulness, problems raised and solved, and avoidance of redundancy etc. rather than settling this question immediately on the basis of conservativism or familiarity bias. For, if taking conditional logical possibility as a primitive is favored on the former grounds, then (I take it) using it as a primitive when formalizing set theory is (at least for our current foundational purposes) appropriate and acceptable. Also recall that my agenda in this book is only to provide a more satisfying foundation for set theory, support the potentialist response to set theoretic paradoxes, and clarify the relationship of math to logic not to establish the metaphysical/cognitive triviality of mathematics or nominalism about mathematical objects (I accept the existence of various mathematical objects). So I don't need or mean to presume that the notion of conditional logical possibility is an epistemic or metaphysical free lunch (in Part II I'll treat it as a substantive part of fundamental ideology). I don't even need to assume that conditional logical possibility facts are ontologically innocent, although (as I'll note in §4.3.1) I think the nominalist can make a decent case for this.

4.1 Motivation

As modal structuralists like Hellman have observed, mathematicians are unconcerned with questions about the nature and essence of particular objects. They don't care whether the number '1' refers to the set $\{\{\{\}\}\}$ or the set $\{\{\},\{\{\}\}\}$ or Julius Caesar, only that whatever objects the predicate 'natural number' applies to have a certain structure (under whatever relations are expressed by the terms 'successor', '+' '·' etc.). And any copy of this structure (whether formed of sets or emperors) is, in some sense, equally relevant to number theory[2]. Considering any objects under any relations of the right arity will do, provided the right pattern in how these relations apply is instantiated.

However, developing potentialist set theory requires both powerful logical vocabulary and some way to compare logically possible structures, e.g., determine when one initial segment of the cumulative hierarchy extends another. Hellman turns to quantifying in to solve this problem, but I think we can instead extend the insight that mathematics is fundamentally concerned with structure rather than identity to achieve the same goal.

I'll suggest it suffices to reconstruct potentialist set theory to consider what's possible given the *pattern* of how some relations (instantiating some math-

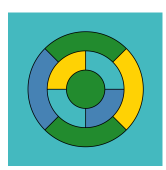
[2]So, for example, suppose you take some strokes that form an instance of the natural number structure under 'to the right of' and erase one stroke and then rewrite a new stroke in the same place, (so that the patterns of how the relations 'stroke' and 'to the right of' apply is preserved but the objects are different). Then you have another copy of the natural number structure and (in a way) nothing that matters has changed. Similarly the particular relations under which objects form a copy of the natural number structure don't matter (turning your stroke sequence on its side and changing each stroke to an exclamation mark so that you now have an omega sequence of exclamation marks under 'below' produces something equally relevant). Any relations of the right arity will do.

ematically relevant structure) apply. To motivate this, note that it doesn't matter to Putnam's potentialist set theory whether some particular objects forming an iterative hierarchy structure (under some relation like 'there is an arrow pointing from... to ...') could *continue to exist* while this structure is supplemented by additional objects so to form an extending iterative hierarchy (vs. whether this is blocked by the metaphysical or logical shyness of the objects as per §3.6). It's quite sufficient for potentialist purposes that the *structure* of how the relation 'there is an arrow pointing from... to ...' applies to these objects to be preserved, while objects forming a suitable extended hierarchy (under some other relation) are added.

4.2 Conditional Logical Possibility

4.2.1 Introducing Conditional Logical Possibility

To see something like the notion of conditional logical possibility (aka logical possibility given structural facts about how some relations apply) arises in natural language, consider claims that some map isn't three colorable.



When you say a map isn't three colorable, you don't just mean that it would be physically or metaphysically impossible for the map to be three colored (without some change in the extensions of 'country on the map' and 'adjacent to'). Rather, you are saying something stronger, which we might make explicit by saying that it's *logically impossible* given the *structural facts* about (aka pattern of) how the relations 'country' and 'adjacent to' apply for the map to be three colored. This means two things.

First, the *mere pattern* of how the relations 'country' and 'adjacent' apply (rather than any special features of the objects in question) suffice to block three colorability. For instance, if wars and revolutions change country names and shift national boundaries but *don't* change adjacency facts then this new map is three colorable only if the old one was, as it's the structure of how the relations adjacent and country apply which determines if the map is three-colorable.

Second, this pattern of how the relations 'country' and 'adjacent' apply makes three coloring *logically* (as opposed to merely physically or metaphysically) impossible, i.e., it blocks three colouring in virtue of completely general, subject matter neutral, laws that treat all relations of the same arity alike. Thus, it's equally impossible for the map to be three scented or three textured either. And if any other relations (e.g., 'city' and has a 'has a direct flight to') instantiated the same pattern, then they wouldn't/couldn't be three colored/textured etc. either.

The notion of conditional possibility ($\Diamond_{...}$) generalizes the notion of logical possibility ($\Diamond$) in a way that lets us naturally express claims like the three colorability statement above. The subscript will specify certain relations – in this case 'is a country' and 'is adjacent to' – whose pattern of application we want to hold fixed. So, as will become clear in a moment, we can write the non-three-colorability claim above as follows:

Non-Three-Colourability $\neg\Diamond_{\text{adjacent,country}}$ Each country is either yellow, green or blue and no two adjacent countries are both yellow, both blue or both green.

I will read this as meaning, "It's not logically possible, given the structural facts about how 'adjacent' and 'country' apply, that: each country is either yellow, green or blue and no two adjacent countries are the same color."

As we saw in §3.5, there's independent reason to accept a primitive modal notion of logical possibility, interdefinable with validity. If you accept this notion of logical possibility, it seems only natural to be able to make sense of restricting that notion to the scenarios which preserves the structure of how some relations apply. To further precisify what I mean, consider a statement like the following.

Crowded Cats Given what cats and basket there are, it is logically impossible that each cat is sleeping in a different basket.

If we take logical possibility to mean logical possibility simipliciter, this sentence must be false. However, it also has an intuitive reading which on which it could be true. One might express the latter by saying 'Cathood and baskethood apply in a way that ensures that (as a matter of mere logic and combinatorics) it can't be that each cat is sleeping in a different basket'. A moment's thought will reveal that (on this reading) the above sentence is true if and only if there are more cats than baskets.

As we saw above, I will express such claims about conditional logical possibility using an operator $\Diamond_{(...)}(...)$. This conditional logical possibility operator takes a sentence ϕ and a finite (potentially empty) list of relation symbols $R_1, \ldots, R_n$ and produces a sentence $\Diamond_{R_1,\ldots,R_n}\phi$ which says that it is logically possible for ϕ to be true, without any change to (structural facts about) how the relations $R_1, \ldots R_n$ apply. But for ease of reading, I will sink the specification of relevant relations into the subscript as follows: $\Diamond_{R_1 \ldots R_n}\phi$. So I'll write the claim about cats and baskets above as follows.

> **Crowded Cats**: $\neg\Diamond_{cat,basket}$ [Each cat slept in a different basket.]

Now let me specify three things about how this notion of conditional logical possibility is to be understood.

The first concerns how conditional logical possibility relates to logical possibility simpliciter. We saw that claims about logical possibility simpliciter ($\Diamond$) concern what's possible if we let both the size of the domain of discourse and the application of relations to that domain vary with complete freedom. In contrast, claims about conditional logical possibility ($\Diamond_{R_1,\ldots R_n}$) concern what's logical possible if we hold fixed the structural facts about how some relations $R_1, \ldots, R_n$ apply (while still letting the size of the domain extending this structure and the application of other relations vary freely).

Second, what does it mean to 'hold the (structural facts) about how some relations apply fixed'? In line with the motivating case above, we should note that keeping the structural facts about how some relations apply fixed doesn't mean preserving these relations' extensions (the particular objects they apply to/relate). Rather it means preserving the pattern of how all these relations apply. So, for example, metaphysically possible scenarios where one cat dies early and one kitten is born early will count as preserving the *structural facts* about what cats and baskets there are (i.e., the pattern formed by how cathood and baskethood apply). And preserving the structural facts about how cat(·) and basket(·) will require preserving: the number of cats, the number of baskets and the number of things (0) that are both cats and baskets. In more familiar Platonistic language, we might say it means holding the extensions (where these are n-tuples for n-ary relations) of these relations fixed *up to isomorphism*.

To bring out the difference between preserving structure and preserving objects at issue, note that I can suspend judgement (or deny that there's a legitimate question) about which properties Nixon had essentially (politician, human, liar, man) while accepting and evaluating claims about what's metaphysically or logically possible given the structure of how certain properties and relations (e.g., 'reports to' and 'is a politician') that are actually satisfied by Nixon and his cronies apply.

Third, note that I don't take structure preservation to require holding fixed the whole size of the universe. The structure which $\Diamond_{R_1,\ldots R_n}$ claims hold fixed is the structure formed by the objects which at least one of the relations $R_1, \ldots R_n$ apply to, considered under the relations $R_1, \ldots R_n$. In this case that means considering the structure of the cats and baskets under the relations cat(·) or basket(·)) [3].

To motivate this way of thinking about what it takes to preserve/agree on structural facts about some list of relations, consider when we'd say two different interpretations of some person's language agree on the *structure* of the natural numbers (under successor). Two interpretations will agree on the structure of the natural numbers if they both take 'number' and 'successor' to apply to some ω sequence,– even if they disagree about the total size of the universe or whether Julius Caesar or the empty set are identical to any numbers etc[4] My understanding of what it takes to keep structural facts fixed generalizes this way of thinking about of what's required to preserve the natural number structure (the structure of objects under the relations 'natural number' and 'successor').

4.2.2 Clarifications and Comparisons

Shapiro On Structures

One can further explain and motivate my the notion of conditional (i.e. structure preserving) logical possibility by relating it to Stewart Shapiro's notion of structures in [109]. Then he says that a structures are 'the abstract form' of a system, which we get by "highlighting the interrelationships among the objects, and ignoring any features of them that do not affect how

[3] Speaking metaphorically, we want to consider logically possible scenarios where there's a bijection f between the set of objects which at least one of the relations $R_1 \dots R_n$ apply to in the actual world, and the set of objects these relations apply to in that logically possible situation, such that $R(x, y)$ iff $R(f(x), f(y))$.

[4] Or consider the way that a Platonist would say the structure of the natural numbers is fixed necessarily and will always remain the same, even if the total size of the universe can be changed by the creation or destruction of physical objects or changes to the structure of space etc.

they relate to other objects in the system." Thus, for example, "The natural-number structure is exemplified by the strings on a finite alphabet in lexical order, an infinite sequence of strokes, an infinite sequence of distinct moments of time, and so on." And adding or subtracting objects to the world outside of a given system, will make no difference to which structure that system instantiates.

Although I mean to propose them as conceptual primitives, one can (roughly) explain my notion of conditional logical possibility (aka structure preserving logical possibility) in terms of Shapiro's notions as follows:

It is logically possible, given the $R_1 \ldots R_n$ facts, that ϕ (i.e., $\Diamond_{R_1 \ldots R_n}$ iff some logically possible scenario makes ϕ true while holding fixed what structure (in Shapiro's sense) the system formed by the objects related by $R_1 \ldots R_n$ (considered under the relations $R_1 \ldots R_n$) instantiates.

Comparison to Logically Possible Worlds

Given my readers' presumed prior familiarity with set theory and metaphysics, it may help indicate the modal notion I have in mind to relate conditional logical possibility facts to common ideas about set theory and possible worlds. However, it should be noted that this comparison is made purely for expository efficiency. I'm putting conditional logical possibility forward as a conceptual and metaphysical primitive which we *could* learn by immersion, in the same way we learn 'set' and '$\in$'.

If we could talk about functions between (the objects in) different logically possible worlds, then we could specify what it takes to hold the structural facts about how some relation (say, 'admires()') applies fixed, in terms of isomorphisms as follows.

A world w_2 counts as holding fixed the structural facts about how 'admires' applies in w_1 iff the objects related by admiration w_1 are isomorphic to those related by admiration in w_2 (you can map one collection of objects to the other in a way that's 1-1 and respects admiration).

> A logically possible world w_2 counts as holding fixed the structural facts about how admires() applies in w_1 iff some function f bijectively maps the objects which either admire or are admired in w_1 to the objects which either admire or are admired in w_2, so that for all objects x and y in w_1 which either admire or are admired in w_1, we have x admires y iff $f(x)$ admires $f(y)$[5]

We will also see that facts about potentialist set theory can be mimiced by talk about models in set theory with ur-elements, in §D below.

4.2.3 Nested Logical Possibility Claims

If we accept the notion of conditional logical possibility, we can also make *nested* logical possibility claims. That is, we can make claims about the logical possibility of scenarios which are themselves described in terms of logical possibility. So, for example, I could say that it would be logically possible for the Crowded Cats claim above to be true.

[5]And more generally, a logically possible world w_2 preserves the structural facts about how relations $R_1, \ldots, R_n$ (say admires() and cat()) apply iff some function f bijectively maps the objects which $R_1, \ldots, R_n$ apply to in w_1 (i.e. those things which are either cats or admire something or are admired by something) to the objects which $R_1, \ldots, R_n$ apply to in w_2 in a way that respects all these relations.

Possibly Crowded Cats: $\Diamond(\neg\Diamond_{cat,basket}$[Each cat slept in a different basket.])

When evaluating such sentences, I think we should hold fixed the meaning of the conditional logical possibility operator (i.e., treat it as a piece of logical vocabulary). And I take the above sentence, $\Diamond$(C$\wedge$B), to a truth because (reading from the outside in):

- It is logically possible (holding fixed nothing) that there are 4 cats and 3 baskets.

- Relative to the logically possible scenario where there are 4 cats and 3 baskets, it is not logically possible (given what cats and baskets there are), that each cat slept in a basket and no two cats slept in the same basket.

Note that in a nested claim with the form $(\Diamond\neg\Diamond_R\psi)$, the subscript freezes the facts about how the relation R applies in the scenario being considered, which may *not* be the state of affairs in the actual world[6].

Based on these kind of examples, I take logical possibility sentences of the form $\Diamond_{R_1,\ldots,R_n}\phi$ to be meaningful, even in cases where ϕ is itself a sentence which makes appeal to facts about logical possibility. And I will work in a formal language $\mathscr{L}_\Diamond$, which I will call the language of logical possibility,

[6]So, for example, $\Diamond$CATS expresses a metaphysically necessary truth. For, whatever the actual world is like, it will always be logically possible for there to be, say, 3 cats and 2 baskets. And any such scenario is one in which it is logically necessary (holding fixed the structural facts about what cats and baskets there are) that: if each cat slept in a basket then multiple cats slept in the same basket. So it is metaphysically necessary that $\Diamond$CATS even if the actual world contains more baskets than cats.

that allows such claims[7]. However, as foreshadowed above, the language of logical possibility will not include sentences which quantify in to the $\Diamond$ of logical possibility, i.e. sentences of the form $(\exists x)\Diamond\phi(x)$.

4.2.4 Comparison to Set Theory With Ur-Elements

We can use the familiar background of actualist set theory *mimic* intended truth conditions for statements in the a language containing the logical possibility operator $\Diamond$ alongside usual first order logical vocabulary (where distinct relation symbols R_1 and R_2 always express distinct relations) as follows.

> A formula ψ is true relative to a model $\mathscr{M}$ ($\mathscr{M} \models \psi$) and an assignment ρ which takes the free variables in ψ to elements in the domain of $\mathscr{M}$[8] just if:
>
> - $\psi = R^k_n(x_1 \ldots x_k)$ and $\mathscr{M} \models R^k_n(\rho(x_1), \ldots, \rho(x_k))$.
> - $\psi = x = y$ and $\rho(x) = \rho(y)$.
> - $\psi = \neg\phi$ and ϕ is not true relative to $\mathscr{M}, \rho$.
> - $\psi = \phi \wedge \psi$ and both ϕ and ψ are true relative to $\mathscr{M}, \rho$.
> - $\psi = \phi \vee \psi$ and either ϕ or ψ are true relative to $\mathscr{M}, \rho$.

[7] To describe this language more explicitly, fix some infinite collection of variables and relation symbols of every arity together with $\perp$ and define the language of logical possibility to be the smallest language built from these variables using these relation symbols and equality closed under applications of the normal first order connectives and quantifiers and $\Diamond_{\ldots}$ (where $\Diamond_{\ldots}$ expressions can only be applied to sentences (so there is no quantifying in). We will also use $\Box_{\ldots}$ in our sentences but regard it as an abbreviation for $\neg\Diamond_{\ldots}\neg$.

[8] Specifically: a partial function ρ from the collection of variables in the language of logical possibility to objects in $\mathscr{M}$, such that the domain of ρ is finite and includes (at least) all free variables in ψ

- $\psi = \exists x \phi(x)$ and there is an assignment ρ' which extends ρ by assigning a value to an additional variable v not in ϕ and $\phi[x/v]$ is true relative to $\mathscr{M}, \rho'$[9].

- $\psi = \Diamond_{R_1 \ldots R_n} \phi$ and there is another model $\mathscr{M}'$ which assigns the same tuples to the extensions of $R_1 \ldots R_n$ as $\mathscr{M}$ and $\mathscr{M}' \models \phi$.[10]

Note that this means that $\bot$ is not true relative to any model $\mathscr{M}$ and assignment ρ.

If we ignore the possibility of sentences which demand something coherent but fail to have set models because their truth would require the existence of too many objects, we could then characterize logical possibility as follows:

> **Set Theoretic Approximation**: A sentence in the language of logical possibility is true (on some interpretation of the quantifier and atomic relation symbols of the language of logical possibility) iff it is true relative to a set theoretic model whose domain and extensions for atomic relations captures what objects there are and how these atomic relations actually apply (according to this interpretation) and the empty assignment function ρ.

[9] As usual (?) $\phi[x/v]$ substitutes v for x everywhere where x occurs free in ϕ

[10] As usual, I am taking $\Box$ to abbreviate $\neg\Diamond\neg$

4.3 Advantages

Given that we can state Putnamian potentialist set theory in the terms Hellman provides, why should we consider adopting an unfamiliar notion like structure preserving logical possibility (aka conditional logical possibility)? Although it might seem that taking the notion of structure preservation as a primitive is a very substantial assumption, I think the cost is more than offset by the benefits it provides.

First, of course, it lets you do potentialist set theory without incurring controversial commitments to object essences and cross-world identity facts as discussed in section 3.5. So it lets you assert axioms which can be accepted by those who share Quine's doubts about whether quantifying in is meaningful and avoid Linnebo's shyness worry [11].

Second, even if you don't object to the metaphysical primitives required to make sense of quantifying-in (and feel that most relevant disagreement could be revealed to be mere verbal disputes by more carefully evoking Hellman's intended reading of quantifying-in) stating axioms in terms of conditional logical possibility helps us articulate principles that can be widely and easily recognized as true. Of course, this is not to say that people can't philo-

[11] Note that one might well accept that there are definite facts about metaphyscial or logical possibility such as could be 'coded' by set theoretic models specifying the size of the domain and extensions for properties, while not thinking there are meaningful (or non-context relative) facts about essences. I can specify the facts about what metaphysically possible worlds there are in terms of how many objects exist in each and how all properties apply (hence pining down facts about what's conditionally logically possible with respect to each possible world w and list of relation $R_1 \ldots R_n$), without telling you anything about essences or counterpart hood relations which would let you determine facts about what's de re metaphysically possible for a given individual. Thus one might well think it's meaningful to ask about whether it's structure preservingly possible that ϕ without asking whether certain particular objects could exist in a world where ϕ.

sophically disagree about the meaning of the conditional logical possibility operator[12], or that disagreement over mathematical axioms is completely impossible. However, there aren't multiple widely held views about the nature of logical possibility which would assign different truth values to commonly used sentences.

Third, there's a practical benefit to using the conditional logical possibility operator rather than quantifying because it in (in effect) cleaves good reasoning about logical possibility into two parts.

- In one part we use standard first-order logic to reason about a given logically possible scenario/what an arbitrary logically possible scenario must be like.

- In another part we use special modal-structural principles to establish which scenarios are logically possible, and 'transfer' facts about one scenario to another.

This helps us avoid the potentially confusing and hard to survey interactions between modal principles and free variables that we see in examples like the proof of the converse Barcan Marcus formula. This is an especially important property for foundational axioms to have as their truth should be evident.

[12] There are plenty of ways of disagreeing about how the intuitive notion of conditional logical possibility should be cashed out e.g., disagreements about whether logical possibility simpliciter should be understood in terms of possible interpretations for words transfer to this case.

4.3.1 Ontology and Conceptual Primitives

Finally, let me end with a brief remark about the impact of this notion of conditional logical possibility on debates about ontology.

In this book I will argue that (at least for a certain foundational project) we should make one choice of primitive, while another has been historically familiar. I'll suggest that philosophy of set theory will go better in certain ways (e.g., we can do a better job avoiding intuitive paradoxes and fleshing out/explaining why apparently good mathematical arguments are justified and correct) if set theory formulated potentialistically, using the conditional logical possibility operator indicated above. [13]. I ask that readers evaluate this choice of primitives on the basis of philosophical fruitfulness, problems raised and solved, and avoidance of redundancy, rather than by familiarity bias. If taking conditional logical possibility as a primitive is favored on the former grounds, then (I take it) using it as a primitive when formalizing set theory is (in some sense) appropriate and acceptable.

Some readers may fear that the notion above is, or enables, cheating at the project of ontology. However, it should be noted that my aim is not to defend any kind of materialism or nominalism (I'll ultimately argue for the existence of some pure mathematical objects) or argue that facts about set theory are somehow metaphysically or epistemically trivial (or in any other sense a 'free lunch'). Accordingly I don't mean to presume that facts about logical possibility are cognitively or metaphysically trivial – or even ontologically

[13] See chapter 11 for much more detail about this.

innocent[14]. For example, I take the concept of logical possibility to be a significant part of fundamental ideology (something that, e.g., should be counted when evaluating the metaphysical parsimony of any theory that employs it).

But how does this argument for specific practical benefits bear on ontology? Some readers might indeed treat my arguments (if they are successful) as suggesting the conditional logical possibility operator is among the Siderian ideological fundamentalia, and hence that should formalize our best theory using it for the purposes of applying Quine's criterion and wind up being nominalists about set theory. And they might defend this stance as follows:

> **Possible Argument for Ontological Innocence of Conditional Logical Possibility** Admittedly it may be possible to think about conditional logical possibility in a reifying way (e.g. in terms of Shapiro's structures or isomorphisms between logically possible worlds). However, such reifying re-interpretations are no more intrinsically clear or acceptable than the modal way of thinking about these facts that I've advocated. And I take the reflections motivating accepting any potentialism to show that just because you can reify a notion doesn't mean you should. When we see that mathematically/inferentially similar work can be done by either adding to our ontology or ideology, we shouldn't always assume that adding to our ontology is the right way to go.

[14] I will consider what a nominalist who thinks logical possibility facts are ontologically innocent can say about indespensability worries in Part II

> Taking claims about logical possibility to be ontologically innocent doesn't violate the *letter* of Quine's criterion for ontological commitment (no quantification over anything other than countries is involved). And I think any sense in which it could be said to violate the *spirit* of Quine's criterion involves the kind of unjustified presumption in favor of expanding ontology rather than ideology, just criticised[15].

However nothing in the main project of this book (neither the justification for the ZFC axioms in Part II nor the development of neo-carnapian realism about moth mathematical objects and account of applied mathematics in Part III) depends on accepting the above line of thought.

[15] Admittedly more austere philosophers may reject the whole idea that we can find the 'true' conceptual primitives (the ideology of a Siderean[112] maximally joint carving language) by considerations of problem solving, unification, fruitfulness etc., and then use these notions when applying Quine's criterion. They may say: the only meaning ontological questions have is as questions of what we quantify over when stating our theories in a specific logical language (FOL) which doesn't happen to include the conditional logical possibility operator. So if we try to introduce a new logical operator, like my $\Diamond_{\ldots}$, to formulate mathematical claims modally, we're pulling the rug out from under ourselves: stepping outside of the realm of ontology all together.

I take this to be a minority position, insofar as it would equally well immediately rule out attempts to formulate theories using a metaphysical or logical possibility (simplicier) operator for the purposes of applying Quine's criterion and assessing ontological commitments as *nonsensical* — something which I take most metaphysicians who think such views are wrong are not willing to do.

However, to such readers I would say the following. In addition to the classic Quinean project of seeing what existential claims you can prove from your best theory while speaking FOL, there's a similarly formally constrained'post Quineian ontology' project of seeing what existential claims you can derive from formalizing theories using the FOL connectives and the $\mathscr{L}_{\Diamond}$. From the metaphysics deflating point of view, there's no lessening of metaphysical insight in going from one project to the other. And, I submit that, there are some practical advantages.

4.4 Looking Forward

In chapter 14 I will show how to use (nested) conditional logical possibility claims to state Puntamian potentialist paraphrases of set theory which avoids quantifying in (and plural and second order quantification). My story will relate to Putnam and Hellman's formulations as follows.

Unlike Putnam my paraphrases will invoke a notion of logical possibility specifically (so metaphysical limits on the cardinality of objects are irrelevant), and I will develop potentialism without claiming that some actualist perspective on set theory is equally good. But Unlike Hellman I will employ specific non-mathematical relations of suitable arities (e.g., 'pencil point' and ' an arrow from...to...' or 'angel' and 'admires') to talk about the possible existence of iterative hierarchies structures, rather than second order variables X, f (or pluralities simulating such variables). I will also differ Hellman in considering iterative hierarchies satisfying IHW rather than hierarchies satisfying ZFC_2.

Obviously, unlike both Putnam and Hellman, I'll offer paraphrases for set theory that don't 'quantify in' Rather, I'll use conditional logical possibility operator to express claims about the possibility of extending iterative hierarchy of sets structures, without detour through claims about logical essences[16].

But in fact, I'll further diverge from Putnam and Hellman by eliminating all second order and plural quantification. For it turns out that the con-

[16] And we will see that talk about what's possible while preserving a given choice of objects/positions x, y within such hierarchies can be handled similarly.

ditional logical possibility operator can do the same work as second order quantification in categorically describing the intended structure of iterative hierarchies.

Thus, reformulating Hellman's potentialism with the conditional logical possibility operator turns out to have a further advantage as regards conceptual parsimony.

> Hellman uses: first order logical vocabulary, second order quantification, logical possibility and (to make sense of applied mathematics) something like an actuality operator.
>
> I will use: first order logical vocabulary and the conditional logical possibility operator.

Insofar as a single modal notion turns out to be able to do both of the following jobs in potentialist translation, it is appealing to use it.

- describe the (first order logic transcendent) structure of the partial hierarchies of sets potentialists want to consider
- articulate the sense in which some such structures are possible and could be extended in certain ways.

To put this idea about conceptual parsimony another way, there can seem to be a kind of undesirable conceptual duplication in employing both the $\Diamond$ of logical possibility and a notion of second order quantification while treating these as unrelated primitives (as Hellman does). For, intuitively, there's something in common between the way we consider 'all possibilities' for how some first order predicates could apply when evaluating logical possibility and the way we consider 'all possibilities' for choosing some first order objects from a given collection when considering what second order objects exist.

Perhaps actualists about set theory can straightforwardly explain this similarity. For they can define both notions in terms of what sets exist[17]. In particular they can appeal to the same notion of 'all subsets over a given first order domain' when defining logical possibility in terms of the existence of a model and when cashing out second order quantification in terms of sets existence.

But potentialists cannot do the same. For we potentialists understand set existence in terms of logical/interperetational/whatever possibility, rather than the other way around. So we can't account for the sense of conceptual overlap between the notions of logical possibility and second order quantification by cashing out both notions in terms of set theory. Thus we lose the above benefit and, e.g., Hellman and Linnebo wind up treating logical possibility and second order quantification as just separate conceptual primitives.

Happily, however, we can solve this problem if we embrace the notion of conditional logical possibility, for this single notion can be used to articulate and analyze both claims about logical possibility and second order quantification[18].

[17] Or at least, they can do this if we bracket Field's objection to identifying claims about logical possibility with claims about set theory discussed in 3.5

[18] We have seen how to do this in for the purposes of second order claims needed to

formulate potentialist set theory. See [8] for an argument that we can reformulate second order claims more generally

Chapter 5

Parsonian Potentialism

5.1 Introduction

I will conclude Part I of this book by contrasting the Putnamian school of potentialism to be developed in the rest of this book with a different, Parsonian, approach developed by Parsons[85, 86, 87], Linnebo[71, **?**] Studd[113] and Roberts [105, 104]. In a nutshell, the difference between the Parsonian and Putnamian schools is that Parsonians interpret set theory as talking about what *sets* (as objects with a special kind of essence) could be formed, while Putnamians understand set theory as claims about how structures satisfying an explicit axiomitization for an initial segment of the set-theoretic hierarchy (e.g. ZFC_2 or IHW_2) could be extended.

At first glance, the choice between Parsonian and Putnamian approaches to set theory makes little difference to our foundational project. Advocates of both views have proved that their favored potentialist translations of all theorems of ZFC are provable from certain modal principles, but don't claim prima facie obviousness for (all) these modal principles. Indeed both sides tend to, in effect, take their potentialist translation of replacement as an axiom (sometimes noting that similar assumptions have been made elsewhere [**?**]). However, we'll see that it turns out to be more convenient to adopt a Putnamian framework (at least temporarily) for this justificatory project.

In this chapter, I'll discuss major existing Parsonian proposals, and motivate my choice to use the Putnamian framework. In §5.2 I'll describe the basic structure of Parsonian paraphrases for set theory. In §5.3 I'll contrast social constructivist and interpretationalist Parsonianism. In §5.4.1 I'll discuss the particular versions of interpretationalism developed by Studd and Linnebo. In §5.5 I'll argue that it's acceptable (and foreshadow why it's useful) to work in a Putnamian framework, even if we ultimately want to be Parsonian potentialists. Finally, in §5.6 I'll compare the merits of Putnamian and Parsonian approaches (with special focus on the forms of interpretationalist Parsonianism developed by Linnebo and Studd).

5.2 The Parsonian Approach

In [73] Linnebo explains the contrast between his preferred Parsonian approach to potentialist set theory and the Putnamian potentialism discussed above as follows.

> [On a Parsionian approach to set theory] the idea is not to

> ‘trade in’ one’s mathematical objects in favor of modal claims about possible realizations of structures but rather to locate some modally characterized features in the mathematical objects themselves. The mathematical universe is not ‘flat’. Rather, some of its objects stand in relations of ontological dependence, and the existence of some of its objects is merely potential relative to that of others.
>
> ‘A multiplicity of objects that exist together can constitute a set, but it is not necessary that they do. Given the elements of a set, it is not necessary that the set exists together with them. ... However, the converse does hold and is expressed by the principle that the existence of a set implies that of all its elements.’ (Parsons, 1977, pp. 293–4)[84]
>
> As Parsons emphasizes, this approach can also be used to explicate the influential iterative conception of sets, which tends to be explained by suggestive but loose talk about a ‘process’ of ‘set formation’. It would be better, Parsons claims, to replace this talk of time and construction with ‘the more bloodless language of potentiality and actuality’.

So the Parsonian takes the term ‘set’ to have pre-existing meaning (and facts about the essential nature of sets to do critical work in their theory), while (as we have seen) the term ‘set’ is completely eliminable from the Putnamian’s theory. And Parsonian potentialists take facts about what pure sets exist to be (in some sense of the word) contingent, with the existence of a set requiring the existence of that set's elements, but the overall height of the hierarchy of sets being contingent. Accordingly, Parsonian paraphrases of set theoretic sentences have a similar large-scale structure to Putnamian paraphrases, replacing $\exists$ claims with $\Diamond$claims and $\forall$ claims with $\Box$s. However, they take the relevant notion of possibility to concern what sets could (in some relevant sense) be formed. And Parsonians don't write any description of the iterative hierarchy structure into their potentialist paraphrases. Instead, they take the fact that whatever sets exist must form (part of) an iterative hierarchy to fall out of — and be explained by — facts about the essences of sets and dependency relations between them.

For example, we saw that a Putnamian like Hellman might paraphrase "$(\forall x)(\exists y)(x \in y)$" [1], as follows.

$$\Box(\forall V_1)(\forall x)[x \in V_1 \rightarrow \Diamond(\exists V_2)(\exists y)(y \in V_2 \wedge V_2 \geq V_1, \wedge x \in y)]$$

If we were to fully expand out the notation above, the resulting sentence would only use modal and logical primitives (not including either set or $\in$).

In contrast Parsonians would translate "$(\forall x)(\exists y)(x \in y)$" more simply along the following lines.

$$\Box(\forall x)[\text{set}(x) \rightarrow \Diamond(\exists y)(\text{set}(y) \wedge x \in y)]$$

They'd then appeal to substantive assumptions about set essences and what they entail about the possibility of set formation. For example Linnebo and Studd take the fact that whatever sets have been formed always fit into an

[1]Recall that here we are using quantification over all V_i as shorthand for quantification over all second order objects X, f (or pluralities simulating them) satisfying some axioms like ZFC_2 (in the sense that $ZFC_2[set/X, \in/f]$)

iterative hierarchy to be explained by facts about sets and plurals like the following.

- There are pluralities xx corresponding to (so to speak) all possible ways of choosing some objects that already exist (e.g., some sets that have already been formed)

- Whenever there's a plurality xx of sets, a corresponding set (i.e., a set whose elements are exactly the members of the plurality) could be formed.

- Sets and pluralities have their elements necessarily (so a set can't be formed before its elements have been formed), and sets are extensional (i.e., two sets are identical iff they have the same elements).

Thus we could imagine a Parsonian hierarchy of sets growing as follows (if we knew what forming a set involved). The empty plurality always exists. So an empty set could be formed. Form it. Now there's a plurality xx whose sole member is the empty set, so a set $\{\{\ \}\}$ could be formed. Form that. Now that both these sets exist then so there are four pluralities xx of sets. And two of them correspond to sets we don't already have. So we could form $\{\{\{\ \}\}\}$ and $\{\{\{\ \}\}, \{\}\}$ etc.

Remember, however, there are two readings of set theoretic talk. In specially literal philosophical contexts, like the paragraph above, we can quantify over the sets that literally exist. However, in mathematical contexts, talk which appears to say that certain sets exist is always shorthand for corresponding claims about what sets could be formed.

5.2.1 Which modal notion?

One obvious question (and potential source of problems for the Parsonian) is this: how shall we understand the Parsonian's modal notion $\Diamond$? In what sense *could* there have been a different number of (pure) sets? And how many sets are there really? For example, if we understand talk of possible set formation as making a claim about how one could reconceptualize the world to think in terms of more sets (as Linnebo does), how many sets are mathematicians *currently* thinking and talking in terms of? It's prima facie unclear how the Parsonians can resolve this tension in a principled fashion (especially if mathematical practices is always better understood by interpreting mathematicians as thinking potentialistically).

Parsons [] argues that we can't understand the possibility invoked in Parsonian paraphrases as meaning physical, metaphysical, mathematical or logical possibility as follows. One can't appeal to physical or metaphysical possibility, because the existence of sets isn't physically or metaphysically contingent. Similarly Parsons understands mathematical possibility to mean possibility dropping 'all constraints of a metaphysical nature' and considering only what is 'compatible with the laws of mathematics' (where the latter include facts about what set exist). Thus he also holds that it wouldn't be mathematically possible for there to be a larger/smaller set theoretic universe.

What about logical possibility? Linnebo notes that [72] appeal to "'logical modality in the strict sense'.. is fairly quickly set aside by Parsons, who finds it to be 'either ... an awkward notion generally or not in the end [different] from mathematical modality." Now I take the arguments of §3.5 to show that there is a very natural and appealing notion of logical possibility (interdefinable with validity) that differs from the kind of mathematical possibility Parsons seems to have in mind. However we cannot interpret the ◊ occurring in Parsonian formalization of set theory to mean logical possibility in this sense. For key claims that the Parsonian wants to say are necessary (e.g., the fact that the sets are extensional) aren't logically necessary[2].

So what modal notion should the Parsonian invoke?

5.3 Constructivist vs. Interpretationalist Options

5.3.1 The Constructivist Option

One option, suggested by talk about generating sets at face value, would be say that sets are literally brought into being — perhaps by some act of social construction, like that which creates contracts and corporations[3]. For example, one might say that adopting an acceptable new axiom of set theory suffices to socially construct or extend the hierarchy of sets up to a sufficient height to satisfy all of ones (now expanded) set theoretic axioms.

Now what about Parsons' point that sets exist metaphysically necessarily? One way of developing this social constructivist approach would be to bite the bullet and reject the above idea. One might propose an error theory

[2] Kit Fine makes a version of this point in [].

[3] See [24] for a proposal that mathematical objects are socially constructed in the same way as marriages and corporations. But note that Cole doesn't doesn't take mathematical objects have the temporal features needed to drive the potentialist story.

about why we falsely think the sets exist necessarily along the following lines. It sounds odd to deny that sets exist necessarily and timelessly because (as noted above) in all normal mathematical contexts apparent claims about set existence really express modal claims (as per the Parsonian paraphrase strategy). And the potentialist paraphrase of the claim that some sets exist really is a timeless necessary truth.

Alternately, one might reconcile the idea of layers of the set theoretic hierarchy being socially constructed with the idea that all mathematical objects are metaphysically necessary and timeless, by drawing on some ideas from Cole[25] and Searle[106] about social construction and the possibility of decisions (about when a company came to exist, or when a player first qualified as on the injured list) taking effect retroactively[4].

However, I take it that significant work would be needed either of the above positions. So, it's not surprising that existing Parsonians tend to take a different approach.

5.3.2 Fine on Situational vs. Interpretational Modality

In [42, 41] Kit Fine proposes a notion of interpretational possibility which has been taken up by the two most developed versions of Parsonian set theory in the current literature.

Fine introduces the notion of interpretational possibility by a kind of idealization on claims about how it is (physically or metaphysically) possible to

[4]Both have suggested that objects which are contingently socially constructed at a certain time (e.g., human rights constructed by a court) might nonetheless be necessary and exist eternally.

reinterpret at given speaker. He suggests that certain acts of reinterpreting a speaker (e.g., by taking their quantifiers to range over an additional layer of sets) witness, but are not necessary for, the interpretational possibility of there being more sets. For example, Fine writes

> [I]t seems clear that there is a notion of [of possibility] such that the possible existence of a broader interpretation is ... sufficient to show that [a] given narrower interpretation is not absolutely unrestricted. For suppose someone proposes an interpretation of the quantifier and I then attempt to do a 'Russell' on him. Everyone can agree that if I succeed in coming up with a broader interpretation, then this shows the original interpretation not to have been absolutely unrestricted. Suppose now that no one in fact does do a Russell on him. Does that mean that his interpretation was unrestricted after all? Clearly not. All that matters is that the interpretation should be possible. But the relevant notion of possibility is then the one we were after; it bears directly on the issue of unrestricted quantification, without regard for the empirical vicissitudes of actual interpretation.[42]

Fine contrasts the notion of interpretational possibility with 'circumstantial' modalities like physical and metaphysical possibility. Interpretational possibilities are supposed to be (as Fine puts it) a kind of, "possibilities for the actual world", rather than "possible alternatives to the actual world." Many different things are interpretationally possible relative to the actual world (as, perhaps, witnessed by the fact that we could interpret someone as speaking with implicit quantifier restrictions and talking about more or fewer of the objects we are currently quantifying over). Contingent differences to what the world is actually like are supposed to make no difference to interpretational possibility. Fine writes, "Circumstance could have been different; Bush might never have been President; or many unborn children might have been born. But all such variation in the circumstances is irrelevant to what is or is not [interpretationally] possible."[42]

Accordingly, there is no conflict between saying it's metaphysically necessary that the hierarchy of sets stops at a certain height and that it's is interpretationaly possible for it to have a different height. And interpretatonalists Parsonians see no tension between understanding possible set formation in terms of interpretational possibility and accepting the intuition that sets exist necessarily.

Fine ultimately rejects understanding mathematics in terms of interpretational possibility, but (as noted above) both Linnebo and Studd invoke use his notion of interpretational possibility to develop their versions of Parsonian set-theoretic potentialism.

5.4 Linnebo and Studd

In this section I'll discuss how versions of Fine's notion of intepretational possibility have been used by Linnebo and Studd to develop Parsonian potentialism. I'll also foreshadow some worries (to be raised below) about the principledness and attractiveness of these versions of interpretational possibility as a choice of conceptual primitive.

5.4.1 Linnebo's Interpretational Possibility

In [75] Linnebo develops a version of Parsonian potentialist set theory which invokes the above notion of interpretational possibility and connects it very directly to Frege's notion of abstraction principles. He develops a version of Parsonian potentialism within a larger account of how we can shift our language (to conceptualize the actual world in terms of more objects) by adopting abstraction principles. He suggests that some objects are 'thin' (with respect to some other objects), in the sense that we can come to know things about the former thin objects by introducing abstraction principles that specify indentity conditions for them by appeal to the objects they are thin with respect to. For example, in Frege's classic case, if you are already talking about lines, you can start talking in terms of the abstract objects we call 'directions', by stipulating that two lines have the same direction iff they are parallel.

Accordingly, we can interpret talk of 'forming' new objects as making claims about how one could (re)conceptualize the world as containing additional objects. Linnebo writes that he will take, "modal operators $\Box$ and $\Diamond$ to describe how the interpretation of the language can be shifted-and the domain expanded-as a result of abstraction."[75] $\Diamond\phi$ is supposed to be true if you could make ϕ true via some well-ordered sequence of acts of reconceptualizing the world via adopting abstraction principles, whether or not it would be metaphysically possible for anyone to make such a sequence of abstractions.

Note that the adoption of such abstraction principles doesn't bring anything into being – whether it be a physical object or an abstract object. Rather it involves "reconceptualizing" the world via the adoption of abstraction principles. Also note that Linnebo only considers reonceptualizations which recognize more objects not ones which remove objects we currently recognize. Since his notion of possibility only allows the world to grow, it doesn't satisfy S5 (unlike logical possibility)[5] and Linnebo accepts the converse Barcan Marcus formula as true with regard to interpretational possibility.

Importantly Linnebo appeals to a notion of **dynamic** abstraction, which lets one expand the application of some previously understood notion by adopting an abstraction principle. One can, in effect, introduce a predicate 'Old()' that applies to all of ones old objects and then adopt abstraction principles that say that for every plurality of old sets there's 'set' collecting exactly these objects. We might think of the above abstraction principle as saying, 'I'll continue to refer to all these old objects and start accepting certain abstraction sentences implying there are new ones which relate to the old objects in a certain way.' This has the important effect that repeatedly adopting (syntactically) the same abstraction principle can lead you to talk in terms of longer and longer hierarchies of sets.

Finally, Linnebo holds you can only start thinking in terms of a set if you are already (or simultaneously start) thinking in terms of its elements (paradigmatically a set is introduced by adopting abstraction principles that say that threre's a set collecting every plurality of old sets). This gives us the

[5] Speaking in terms of Kripke models, when it comes to interpretational possibility only worlds that preserve or add to the objects existing in a world w_0 are accessible from w_0.

dependence of sets on their elements referenced in §5.2 above.

Now, like Hellman, Linnebo shows that we can justify the use of the ZFC axioms from certain modal assumptions – in this case, assumptions about what's interpretationally possible. However, it seems to me that some of the assumptions used in this proof raise an important question about how Linnebo's notion of interpretational possibility is to be understood. In particular, Linnebo appeals to the following maximality principle.

> "Maximality: At every stage, all the entities that can be introduced are in fact introduced."

But this principle seems very implausible on the intuitive understanding of interpretational possibility as possibility with respect to how, "the interpretation of the language can be shifted — and the domain expanded — as a result of abstraction" evoked above. For surely we don't introduce all possible abstraction principles at once!

Perhaps one can solve this problem by simply understanding Linnebo's notion of interpretational possibility more narrowly (as concerning how the world could be reconceptualized at some stage of a process that *did* simultaneously introduce all possible abstraction principles at each stage). However, such ad hoc restriction of the intuitive notion of 'what can be got to by introducing abstraction principles' above can make the concept of interpretational possibility seem significantly less principled (and hence less attractive as a choice of theoretical primitive) than logical possibility

5.4.2 Studd on Interpretational Possibility

In [113] develops the notion of interpretational possibility in a way that (I'll suggest) raises a similar concern about whether interpretational possibility is an attractive choice of theoretical primitive.

Interestingly, Studd introduces his notion of interpretational possibility by contrasting it with logical possibility, saying that interpretational possibility is like logical possibility but with a stricter accessibility relation. He then says his notion of interpretational possibility is very similar to Linnebo's. However, where Linnebo talks about interpretational possibilities as corresponding to ways of reconceptualizing the world, Studd talks about 'admissible interpretations' for a certain lexicon. And where Linnebo talks about abstraction principles, Studd talks about successful attempts to liberalize our language (by adopting a new, more liberal, interpretation for the lexicon in our language) by — so to speak — expanding the domain of quantification and adding some of these new objects to the extension of predicates like 'set'.

Studd writes, "The truth of $\Box\phi$ depends on whether the proposition that would be expressed by ϕ under other admissible interpretations of the lexicon is true (in the actual world)" and says the following about admissible interpretations.

> Admissible interpretations result from shifts of interpretation of the kind that a [quantifier] relativist may bring about in her attempt to expand the universe. Such interpretations come with

> a natural ordering: an admissible interpretation j is said to succeed another i iff j results from one or more relativist attempts to admissibly liberalize the interpretation i. In this case we also say that i precedes j.

Studd notes that his concept of 'admissible interpretation' (and hence interpretational possibility) differs from the natural language notion of what we could get our words to mean[6]. Instead he uses Kripke-like models and a principle that these models must satisfy certain monitonicity and stability requirements to convey his concept of admissible interpretations and liberalizations. This amounts to requiring that relevant meaning change events have *at least* the following features: they only introduce new objects to the domain of the quantifiers and the extension of 'set' and 'element' without stopping your quantifiers from ranging over anything you are currently quantifying over (Monotonicity) or changing how 'set' and 'element' apply to these current objects (Stability)[7]

[6]He writes, "In the case of the present version of English, for instance, there's nothing to stop us from attaching new meanings to terms like 'set' and 'element' that are wholly unconnected with their current meanings. The new interpretation could reinterpret these terms to be coextensive with 'sloth' and 'eats' (as the latter terms are presently interpreted). The resulting interpretation is clearly available to us but inadmissible because it fails to meet the Stability constraint. All the same, since this sort of reinterpretation is clearly orthogonal to issues concerning absolute generality, nothing is lost by taking the interpretational modal operators to only generalize over admissible interpretations."

[7]More specifically he explains his meaning metaphorically by providing something like a Kripke a model for his modal notion (with objects called indexes corresponding to specific interpretational possibilities). After conjuring the image of stages in a growing hierarchy of sets with indexes corresponding to particular stages of growth, Studd writes,

> "Less metaphorically, we can helpfully think of the indices [of these Kripke models] as admissible interpretations of the sort that the modality is intended to generalize over, with Monotonicity and Stability serving to constrain the sorts of interpretation the modality generalizes about."

In this model we have a set i of indexes for admissible interpretations i_1 i_2 etc. for "S" and "E" (for 'set' and 'element of')

However these constraints cannot be *all* that's required of admissible interpretations. For example, expansions of a language which add a new object to the extension of 'set' and not the extension of 'element' despite that language already recognizing an empty set (so the extensionality axiom is violated), are presumably not admissible interpretations. And later, when justifying matheamticians' use of the ZFC axioms, Studd makes a plenitude assumption that amounts to saying that whenever you liberalize the meaning of set to 'add' one set, you must thereby add (at least) a full layer of new sets.

Overall Studd says rather little about how he understands interpretational possibility, beyond the points summarized above and some further principles specific to set theory. It's not clear to me that any single unified intuitive notion implies all the constraints on interpretational possibility Studd asserts (or whether Studd even claims to have latched on to such a notion). Thus, I think Studd's notion of interpretational possibility can, like Linnebo's, seems unprincipled and ad hoc in a way that's undesirable when developing a foundations for mathematics (even if it's no problem for the project of

> a model-theoretic (or MT-) interpretation of the non-modal language L_{SU} is a set-structure$\langle M, S, E\rangle$ that supplies a non-empty set M as the universe of discourse, and extensions S and E based on M for the language's two non-logical predicates, the set and element–set predicates (β and ϵ).
>
> ...an MT-hierarchy is an indexed-set of triples $\{\langle M_i, S_i, E_i\rangle : i \in I\}$ each member of which is either an MT-interpretation or the empty interpretation (with M_i non-empty for some $i \in I$), and which meets the ... three conditions [serial well order, monotonicity, and stability]

Monotonicity. "Whenever i and j are indices in I with $i <_I j, M_i$ is a subuniverse of M_j (i.e. $M_i \subseteq M_j$)."

Studd also has a requirement of stability which requires that different admissible interpretations agree on the application of element-of on sets they both acknowledge.

Stablity "Whenever i and j are indices in I, the extensions S_i and S_j and the extensions E_i and E_j agree on their common domain $M_i \cap M_j$" .

defending quantifier relativism which most interests Studd in [113]).

5.5 Which Theory to Choose?

With this picture of the most developed versions of Parsonian set theory in mind, I will now attempt to motivate my choice to work in the Putnamian framework (at least for temporary practical purposes).

I'll argue that Parsonians face some pressure to accept the equipment needed for Putnamian paraphrase (and perhaps the correct truth values of these paraphrases). Then I'll explain (rather abstractly) why working in a Putnamian framework will be convenient for my justificatory project.

5.5.1 Acceptability of Logical Possibility to Parsonians

First note that Parsonians Linnebo and Studd do seem to accept the meaningfulness of logical possibility and seemingly agree that (at least some versions of) Putnamian paraphrases have the correct truth-values. As we saw above, Studd introduces his notion of interpretational possibility by appeal to logical possibility, and he even seems to (in some sense) endorse the adequacy of Hellman's Putnamian potentialism[8]. And Linnebo's criticisms of

[8] In a footnote to his chapter on potentialist set theory, Studd writes the following about Hellman's modal structuralist approach (and gives no later criticism of that view),

> "An alternative is for the relativist to adopt modal-structuralism in the style of Hellman (1989).This permits her to interpret $\Box$ simply as logical necessity. On this view, there is no need for admissible interpretations to satisfy the Stability constraint on the interpretation of the non-logical vocabulary set out below. This is because modal structuralism takes set-theoretic statements to be elliptical for statements in a higher-order modal language, which eliminates occurrences of the set and element–set predicate. See also Hellman ... who applies modal-structuralism to offer a potentialist Zermellian response to the set-theoretic paradoxes."

Hellman-style Putnamian potentialism in the his head-to-head comparison of Putnamian and Parsonian potentialism in [73] are strikingly moderate and don't center on raising doubts about the intelligibility of Hellman's proposal.[9]

I think this apparent willingness to accept the meaningfulness of a notion of logical possibility (and something like the intuitions about it the Putnamian potentialist needs to appeal to) is no accident, but rather flows from something basic about Putnamian paraphrases.

For, note that both Studd and Linnebo both take for granted the notion of well-founded sequences of reconceptualization/liberalization events in developing their concepts of interpretational possibility and potentialist set theory. And, arguably any modal notion that could do the work the Parsonian needs must appeal to a very idealized notion of how some objects could

---

[9] After raising the metaphysical shyness and compossibility worry we discussed in §4.3.1 , Linnebo notes:

> Let me be very clear about my complaints in this section. I am not asserting that meta- physically shy objects are in fact possible or that there might not be some clever way to...circumvent the problems generated by the phenomenon of incompossibles. My point is only that the extra freedom of Putnam's approach, which initially seemed purely advantageous,has the unintended side effect of incurring potentially problematic metaphysical commitments, which are avoided on the Parsons approach.

Linnebo's other points against versions of the Putnamian approach in that paper merely involve correctly pointing out version of some points already discussed above: that it's hard to make sense of Putnam's dual perspective (mathematics being equally well understandable in modal and ontological terms), and that Hellman's requirement that initial segments be models of ZFC_2 seems unmotivated and troublesome. Finally, I've argued that Linnebo's shyness based criticism of Putnam's account can be avoided by reformulating Putnamian proposals using the conditional logical possibility operator as per Chapter 4. Some nominalists might worry about implicit commitment to abstract objects – but Parsonian potentialists who embrace the existence (or at least possibility) of sets will not have that doubt.

be well ordered by a relation. But it's hard to see how one could understand the notion of a well-ordered sequence of language changes, without a background notion of something like logical possibility.

Pressure for the Parsonian to accept a notion of how it would be possible (in some sense that isn't hostage to facts about metaphysical possibility) to have a sequence of set-formation events satisfying the well ordering axioms is clearest if we understand the Parsonian to allow only adding a single layer of sets at each stage. But, even if we allow arbitrarily many sets to be introduced at any stage, the Parsonian would be hard pressed to try and insist that it's enough to consider only well-ordered sequence of reconceptualization/ events of some limited height[10]

Thus, one might argue that the Parsonian already needs to understand all the notions needed for Putnamian potentialism. Indeed one might argue that the Parsonian is already appealing to a Putnamian potentialist picture of the *ordinals* to motivate their story. For, the Parsonian must already take there to be a fact of the matter about what well-ordered sequences are possible in some sense that's obviously meant to be free of any purely physical or even metaphysical limitations. If they are going to accept the meaningfulness of asking if there is a well-ordered sequence with a certain property, it would be unattractive to suggest that such talk of coherence or

[10]For example, suppose that all relevant height increases could be performed by repeating a single set generating ceremony $< \alpha$ many times. In actualist terms this would amount to assuming that the whole hierarchy has cofinality α (something widely regarded as implausible by mathematicians). For, in actualist terms it would imply that some second order function f and perhaps even some first order definable relation ϕ could map α to all the oridinals, contrary to the (the spirit of, and second order formulations of) axiom of replacement! And a similar conclusion follows if menu of different abstraction techniques one can in principle apply is small relative to α.

logical possibility is only meaningful for well-orderings and nothing else[11].

Accordingly, I take it that Parsonian potentialists generally do and plausibly should accept the meaningfulness of logical possibility (together with other tools needed to develop Putnamian set theory).

5.5.2 Putnamain Potentialism and Logical Possibility

Now, one might use the point above to argue that Putnamian potentialism (cashed out in terms of some form of logical possibility) should be favored on grounds of ideological parsimony. If one already has to accept logical possibility, and that suffices to let you reconstruct set theory and answer Burlli Forti paradox in mostly the way the Parsonian wants, why accept any additional modal primitive? I will consider such an argument below.

However, in this section I just want to make the following practical point. If I succeed in providing in justifying the Putnamian potentialist version of the axiom of replacement from principles that seem clearly true, Parsonians can plausibly use this result to (at least somewhat) further justify *their* version of the axiom of replacement by inferring it from the Putnamian version of replacement.

This is fortunate because working in a Putnamian framework (of the kind advocated in previous chapters) turns out to be quite convenient. For my

[11]Note that the second order/plural quantification vocabulary typically used to formulate the claim that a sequence of growth events is well ordered also lets you pin down intended models of the iterative hierarchy up to width. So they need to accept not just the logical possibility of first order facts but (something like) logical possibility claims about a hierarchy of growth/reconceptualization events satisfying the non-first order least element condition of well foundedness.

proposed justification of the axiom of replacement leverages intuitions about logical possibility generally in a way that seems difficult if not impossible to reproduce reasoning solely about possible extensions of the set-theoretic hierarchy.

Specifically, my justification provides something analogous to a 'non-elementary proof'[12] in that by considering the notion of logical possibility generally we can derive results about the specific logical possibility claims used in potentialistic set theory. It justifies the potentialist translations of set theoretic claims (claims about how iterative hierarchies can be extended by other iterative hierarchies) by first proving things about how any hierarchy of sets structure could be extended by certain larger structures that *aren't* iterative hierarchies. But it's difficult to see how to reconstruct such reasoning about extendability via larger logically possible structures (chosen for mathematical convenience alone) working purely within in an interpretationalist Parsonian framework, where growth events seemingly only add objects falling under some currently understood indefinitely extensible concept[13]).

[12] Sometimes the easiest or most illuminating way to prove something about the natural numbers is to consider them within some larger structure like the real numbers. Such proofs are called non-elementary proofs.

[13] For example, the Parsonian might try to mirror the reasoning above by considering the interpretational possibility of a hierarchy of sets existing alongside other (non-set) objects. But note, it seems that there might not be any concepts in our current language whose rich meaning (in the sense specified in §5.6.1 below) allows them to form the kind of structure we want to consider our iterative hierarchy embedded in for out 'nonelementary proof'. It seems implausible that, for every describable way it would be logically possible for some relations $R_1 ... R_n$ to pick out a larger structure it's useful to consider a hierarchy of sets being embedded within, it's interpretationally possible for some relations $R'_1 ... R'_n$ to apply in exactly that way. Recall that Studd and Linnebo take there to be various important facts about the meaning of set and element, which ensure that, e.g., you can't think in terms of multiple sets that have exactly the same elements. And presumably the same applies to other current English language concepts as well. Perhaps we could get around this problem by considering intepretational possibilities corresponding to language

5.6 Which Framework to (Ultimately) Choose?

Now lets turn to the question of whether Putnamian or Parsonian potentialism is ultimately to be preferred. Admittedly, this question is a little ambiguous; you might ask 'which version of potentialism should we prefer *for what purposes*?'. For example, we might ask (in a Sideran pro-metaphysics spirit), which formalization of set theory best reveals the facts about fundamental ontology and ideology that ground the truth of set theoretic claims. Alternately one might ask which theory provides the best Carnapian explication of potentialist set theory along lines developed in Chapter 11. However, my remarks below will motivate favoring Putnamian potentialist set theory in both of these ways, so I won't stress the distinction here[14].

5.6.1 Unappealing Choice of Conceptual Primitives

Main Concerns

One argument against (interpretationalist) Parsoniansim questions the attractiveness of interpretational possibility as a conceptual primitive (as compared to logical possibility). In this section I'll note some ways that inter-

changes that adding *new* atomic predicates and relations to our language.

But this approach is difficult to develop in Linnebo or Studd's system. For Studd's talk about interpretational possibility reflects admissible interpretations of 'the' lexicon (rather than considering a lexicon that could be arbitarily extended). And, while Linnebo's system seems to be more open to introducing new concepts, he says interpretational possibilities correspond to we could start talking in terms of by adopting abstraction principles, and it's not clear that every larger structure which it is useful to reason about initial segments being embedded in (for the purposes of non-elementary proofs as above) can be introduced by stipulations which take the form of abstraction principles.

[14]Note that some traditional reasons for favoring platonistic views over modal perspectives on mathematics don't bear on our choice here. Neither Parsonian nor Putnamian potentialists take set theorists' apparent quantification at face value, and both introduces new modal notions go beyond FOL in analyzing set theoretic claims.

pretational possibility can seem much less principled and clearly/concretely understood than the notion of logical possibility, and hence like a less attractive choice of a conceptual primitive for philosophical analysis.

While, perhaps, not very forceful on their own, such worries take on more force given the point that something very similar to the potentialist solution to set theoretic paradoxes Putnamians want to endorse can be developed using a notion of logical possibility that Parsonians seem to (and seem to have reason to) accept. A key contrast between the Putnamian potentialism developed in this book and the Parsonian potentialism advocated by Studd and Linnebo concerns choice of primitives: ought we analyze set theory using a primitive interpretational possibility operator or to analyze both set theory and (something like) interpretational possibility in terms of logical possibility?

In §5.4.1 and §**??** we already saw some reasons for concern that Linnebo and Studd's notions of interpretational possibility must be arbitrarily restricted/non-joint-carving in ways that makes them a bad choice for a conceptual primitive (if they are to satisfy the various assumptions Linnebo and Studd use to vindicate use of the ZFC axioms). Additionally, Linnebo himself notes a way that interpretational possibility facts reflect arbitrary conventions with respect to 'Julius Caesar problems' about when the objects falling under the concept introduced or liberalized by adopting some abstraction principle (e.g., the number 1) are identical to objects one was previously talking in terms of (e.g. Julius Caesar)[15].

[15] He writes, " ...when we develop our linguistic practices, we have some degree of choice about whether or not to allow categories to overlap. To handle mixed identity statements,

Interpretational possibility facts also reflect seemingly highly indeterminate and/or disputed facts about what I'll call 'rich meanings', in a way that can make the interpretational possibility operator seem like an unattractive choice of primitive. The interpretationalist Parsonian needs to distinguish between acts of neo-Carnpaian language change that merely change application of *the word* 'set' and those which also count talking about more *sets* vs. beginning to use the word 'set' to express some other content. For example, they can't allow that it's interpretationally possible that 'there are two sets with exactly the same elements', although obviously one could change the meanings of English words so the corresponding sentence expressed a truth. And presumably the current meaning of the word 'set' (perhaps together with background linguistic conventions and precedents) is what does this work[16].

we often need conceptual decisions, not just factual discoveries.... When our ancestors first confronted Caesar-style questions [i.e., questions like whether Julius Caesar is identical to the number one], they had a choice which way to go; and this choice played a role in shaping the concepts that they thereby forged. Today we find ourselves in a different situation, since many choices are already implicit in the linguistic practices that we have inherited. Of course, insofar as we are willing to revise these practices, we still have the same choice as our ancestors had. But we face an important additional question not encountered by our pioneering ancestors, namely what conceptual decisions are implicit in our inherited linguistic practices. I shall argue that these practices have by and large legislated against the overlap of categories. But exceptions are certainly possible and very likely even actual."[75]

[16] An analogous argument can be made even if we assume that interpretational possibility must satisfy Studd's Maximality and Stability assumptions. For (if you accept the neo-Carnapian view at all) it's intuitively possible to change your language/thought so as to add a new object to the extension of 'set' and not the extension of 'element' (so, given that we're already talking in terms of empty set, the extensionality axiom begins to express a falsehood). But the interpretationalist can't allow that it's interpretationally possible for there to be two *sets* with exactly the same elements (If the interpretational Parsonian allowed this, their paraphrases of the axiom of extensionality coming out false). Thus, there must be some reason that changing your language use in the way indicated above only says something about how one could change the meaning of the word 'set' and not about what *sets* it would be interpretationally possible for there to be. And presumably the current meaning of the word 'set' is what does this work.

Accordingly I interpretationalist Parsonians seem to be committed to our words like set having 'rich meanings', which specify what's needed to preserve the meaning of a word (to continue talking about *sets*) under some neo-carnapian language change that gets us to start talking in terms of new objects (e.g. by introducing an abstraction principle).

Now I admit that we have *some* shared and correlated intuitions about how the meaning of the terms 'set' and 'element' could be preserved neo-Carnapian language change. However it seems to me that such agreement is limited and vexed in the same ways as agreement on the right way to expand the meaning of your terms for the purposes of engaging with a metaphor.

For example, I take it that most people might agree that saying the leader of a country is its 'head' is reasonable way to preserve/honor the current/literal meaning of the term 'head' in a metaphorical context which invites us to apply human anatomical language to parts or aspects of a country. But this limited agreement doesn't provide (or evidence) shared understanding of a sufficiently precise and concretely grasped notion of metaphorical truth (or possibility) to make the latter concept an attractive choice of primitive when logically regimenting mathematics.

Our intuitions about rich meanings, even in the interpretationalist's key case of the concepts *set* and *element* can seem similarly limited. I take it that most people would agree that if we think in terms of more sets, it's natural to suppose these sets would still have to satisfy extensionality. But, suppose I am currently thinking in terms of certain hierarchy of sets V_α. If I were to adopt an abstraction principle which adds an extra layer of 'sets', would this really be a way of thinking in terms of more *sets* (on my current meaning of the term), as the Parsonian account needs? It seems equally or more natural to me to say that, after making this switch, only the 'sets' up to V_α are sets on my current sense of the term, and when expanding my quantifiers in the way suggested I have instead got the word "set" to express a new concept like 'class'. Note that, e.g., the new sets thus introduced won't satisfy the pairing axiom.

Object Identity under Neo-Carnapian Language Change

A similar point about interpretational possibility facts being seemingly controversial and/or indeterminate *may* arise in connection to judgements about *object* preservation under neo-Carnapian language change. Linnebo and Studd can seem to endorse generally determinate facts about when different sequences of abstraction principles wind up introducing the same object[17]

Yet it can seem implausible that there are, in general, such determinate facts about when adopting one sequence of abstraction principles introduces the same entity you could have introduced by some other sequence of abstraction principles.

For example, are the objects you would have introduced by introducing the

[17] For example, in motivating a certain convergence assumption about interpretational possibility Linnebo writes:

> "This principle ensures that, whenever we have a choice about which entities to introduce, the order in which we choose to proceed is irrelevant. Whichever entity we choose to introduce first, the others can always be introduced later. Unless $\leq$ was convergent, our choice about whether to extend the ontology of w_0 to that of w_1 or that of w_2 would have an enduring effect."

.

concept of 'Turing degrees' by abstraction over computations formalized using Turing machines *literally the same entities as* those introduced via abstraction over computations formalized using general recursive functions? And suppose someone who knows that these two notions of computability are equivalent introduces a concept of 'Turing degrees' via abstraction principles involving an unspecified notion of one set of numbers being 'computable from' another which is equally anchored to both definitions of computability. Is this a way of introducing *literally the same entities* you could have introduced by introducing 'Turing degrees' in one of the two ways mentioned above (or merely some isomorphically structured mathematical objects)?

We do, sometimes, say that people whose mathematical definitions differ slightly from ours can 'know things about' structures like the natural numbers or Turing degrees. But arguably what's required for such claims to be true is highly indeterminate and/or context dependent. For example, in most situations it seems reasonable to describe people who know that some claim ϕ holds for any of the Turing-degree-like mathematical structures introduced by any of the acts of abstraction above as knowing something 'about the Turing degrees'. However, in cases where knowing the connection between Turing degrees and general recursive functions or Turing machines specifically matters, we may draw finer distinctions[18].

And note that many philosophers like McGee[80] find it positively attractive to say that reference for abstract terms like natural number (a paradigmatic

[18] Also it's appealing to say that a pair of people who identify the numbers with different ω-sequences of sets (a la Benaceraff's famous paper[3]) both still 'know things about the numbers'. But one *can't* say this is true in virtue of them literally talking about the same objects/numbers, since the number 3 can't be identical to two different sets.

case of objects Linnebo and others want to say are supposed to be introduced via abstraction principles) is only determinate up to isomorphism. Thus we appear to have another dimension along which it seems that facts about interpretational possibility must be controversial, arbitrary and/or indeterminate.

Veil-lifting Picture

Admittedly, there is a certain picture which might motivate thinking there are principled determinate answers to the questions about rich meaning and the identity of objects introduced by abstraction principles above. However, this picture has other, very unattractive features. I will conclude this subsection by discussing it, although I don't mean to claim Linnebo or Studd would endorse it[19]. My point is only that, unless we take everyone to be unveiling portions of some shared total world (in the sense sketched below), it's unclear what would explain there always being definite facts about which Carnapian language changes preserve the meaning of predicates or wind up introducing the same object.

> **Veil Lifting Picture**: There's a shared total world containing all the different kinds of objects anyone could ever talk in terms of. The meaning of each atomic predicate or relation determines its extension within this total universe. Acts of neo-Carnapian language change 'get you to talk in terms of more' objects by lift-

[19]Both certainly say that they mean interpretational possibility to reflect an expansionary modality rather than mere removing of quantifier restrictions from some fixed universe.

> ing parts of the veil' covering a plentiful universe, expanding the domain of objects your quantifiers ranger over to include more things. You can't ever reach a point at which you couldn't expand your quantifiers further. But this is just because (for some reason) no series of linguistic acts could lift the veil completely, so that your quantifiers would range over absolutely all objects that one could, in principle, talk in terms of.

If we take this picture of neo-Carnapian language change as quantifier restriction lifting seriously, we automatically get the required rich meanings (determinate facts about how each property we currently talk in terms of will apply under quantifier meaning shift) and determinate facts about which acts of re-conceptualization by adopting abstraction principles would introduce the same objects. For example there will, prima facie, be determinate facts about whether people who 'push back the veil' by introducing 'Turing degrees' via abstraction over what general recursive functions are unveiling the same objects.

However endorsing general determinate facts of this kind can seem unattractive (as noted above). And the note that this veil lifting picture adds a presumption of determinacy which goes beyond traditional metaphors for neo-Carnapian language change, on which 'different languages carve up the world into objects in different ways'[20] And in Chapter 19 I will propose a

[20] If I carve up some dough one way, but I could have carved it up another way, we don't seem forced to accept determinate *de re* facts about whether a certain cookie brought into being by one carving could instead have been brought into being by another carving and backing sequence. A metaphysician like Kripke *might* choose to directly endorse such facts. But they might equally well analyze such facts away in terms of claims about contextually relevant counterparts, like Lewis. Or they might choose to reject such questions

more concrete approach to neo-Carnapian language change which also avoids this presumption of determinacy.

One might even fear that adopting this veil lifting undermine the motivations for potentialism discussed in §2.2. For on the veil lifting picture it appears there's some shared ineffable domain of all objects which all neo-Carnapian language shifts correspond to quantifying over portions of; we just can't succeed in forcing our quantifiers to be interpreted as ranging over all this structure. But the intuition driving potentialism in §2.2 was that for *any* actual plurality of objects there could be a larger one, not just that any for plurality of objects *we can get our quantifiers to range over* there could be a larger one. So I think it's desirable to avoid the above veil lifting picture if we can.

So, to summarize, facts about interpretational possibility can seem unprincipled (not joint carving), non-fundamental, disputed, and/or frequently indeterminate in ways that make interpretational possibility an unattractive primitive for mathematical or philosophical analysis.

5.6.2 Ideological Parsimony and Conservatism

A different motivation for favoring Putnamian set theory draws on considerations of ideological parsimony. In (§3.5) and (§5.5) we saw some reasons to think that even Parsonians should accept the Putnamians notion of logical possibility. Accordingly, as noted above, one might think Putnamian potentialism should be favored on grounds of ideological parsimony (let us

---

as meaningless (as Quine does in rejecting quantifying in).

not multiply primative modal notions beyond necessity!) and avoiding revisionary and controversial commitments about other areas of philosophy where possible. If Parsonians must accept Putnamian primitives but not vice versa then parsimony surely favors the Putnamian approach.

Admittedly Parsonians could block this argument if they could show that Putnamians also can't do without Parsonian primitives. Interpretationalist parsonianns might argue that we independently need the notion of interpretational possibility to make sense of neo-Carnapian language change. However 18 I'll argue that this is not the case. We can develop a neo-Carnapian philosophy of language sufficient to do the (legitimate) work of neo-carnapian philosophy of language equally well or better using the conditional logical possibility operator. Notably, I'll suggest that doing this legitimate work doesn't require stating claims about the possibility or impossibility of 'absolute generality'/quantifying over anything in some non-trvial sense (the main thing, outside set theory, Linnebo and Studd use the interpretational possibility operator to do).

In addition to adding a new modal operator, interpetationalist Parsonianism also requires us to make some prima facie unintuitive changes to philosophy of language[21]. Linnebo himself says:

[21] Going further, Linnebo's embrace of something like Dummett's indefinite extensability in the quote above may raise prima facie philosophical puzzles, which could easily be avoided by favoring Putnamian potentialism. For it seems to me that the passage above suggests the following picture. There aren't just different *equally legitimate* ways of 'talking in terms of more sets' (for in this case our concept of a hierarchy would merely not be "precise") but rather we have a precise concept with a kind of inadequacy or internal tension, whereby every language including the concept set is held to be in some way *leaving some things out* so that, e.g., languages that talk in terms of more sets are less inadequate than languages that talk in terms of fewer sets. But such ideas about reality forever transcending language and thought can seem prima facie problematic and, and

> Suppose we have formulated a perfectly precise notion of a star. For any object whatsoever, this notion enables a definitive verdict as to whether or not the object is a star. When this precise intension is applied to the world, reality answers with a determinate extension, namely the plurality of objects that satisfy the intension. And there is nothing unusual about stars in this regard. *In most ordinary empirical cases, a precise intension determines an extension when applied to the world. But in mathematical cases, and other cases involving abstraction, this is no longer so. Here a precise intension often fails to determine an extension.*

Thus, overall, one might argue that the main motivation for accepting interpretational possibility (and the claim that precise intensions don't determine precise extensions above) is to account for set theoretic paradoxes while avoiding arbitrariness intuitions[22]. But, if a Putnamian approach can do the same work without requiring us to add to our fundamental ideology or revise general philosophy of language, considerations of parsimony favor the Putnamian framework.

5.6.3 Double Duty

The Problem

Finally, a third challenge for Parsonians concerns the double duty set talk is supposed to play in Parsonian theories.

---

hence desirable to avoid where possible.

[22]Certainly the case of set theory is the main motivation cited for Dummett's project in [30] and Studd's project of defending quantifier relativism in [113].

Recall that the Putnamian potentialist never employs a predicate 'set' in her logical regimentations of set theoretic talk. Thus she could deny that there is a property of being a set. Or if she does accept such a notion, she can say that it has an empty extension, for the same reason 'philogiston' does (sets are objects hypothesized by the wrong account of what set theory is teaching us about).

In contrast, the Parsonian takes ordinary mathematical usage to have given 'set' a definite meaning — enough that there are definite (non-trivial) facts about how tall a hierarchy of sets actually exists/mathematicians are currently thinking in terms of. But it can seem puzzling how mathematicians' set theoretic talk can do this while simultaneously being best understood in a potentialistic fashion. How can such talk determine (in any principled fashion) facts about what height actualist hierarchy of sets they are supposed to be thinking about?

For interpretationalist Parsonians, the challenge looks like this. In what sense can someone said to be 'thinking in terms of' *any* hierarchy of actualist sets with height α, if their set talk should always be interpreted potentialistically? How many sets are Linnebo and Studd currently thinking in trms of? And what principled grounds are there for the answer to this question?

To press this worry (and further clarify Studd's notion of interpretational possibility), I will now discuss Studd's story about our set theoretic practice might unknowingly get our quantifiers to range over more and more sets. It seems to me that if this story worked it might attractively answer the challenge above : providing a principled account of how many sets mathematicians are currently thinking in terms of. However I will argue that it does not work, so the double duty problem remains.

Studd on Expansion and Actual Set Theory

In Chapter 8 of [113] Studd's sketches a story how people with a set theoretic practice like ours could unknowingly change their quantifier meanings and come to talk in terms of a progressively larger actualist hierarchy of sets.

Studd first considers a situation where people knowingly start talking and thinking in terms of extra sets. Imagine that some people start out speaking a language Q. Then they decide to split off from the main body of Q speakers and develop a new language E, which 'talks in terms of' extra sets.

To do this they adopt certain principles, most importantly the inference schemas for reasoning from claims in the old language Q (indicated below by putting 'Q:' in front of them) to claims in the new language E (indicated below by putting 'E:' in front of them) , and vice versa. .

$(U_E - \{\})$

$$Q : things(vv) \Rightarrow E : thing(\{vv\})$$

$$Q : things(vv), Q : v \prec vv \Rightarrow E : v \in \{vv\}$$

$$Q : things(vv), E : v \in \{vv\} \Rightarrow Q : v \prec vv$$

Intutively these schemas embody the idea that each plurality *vv* of objects quantified over in the old language Q is supposed to form a set in the new language[23]. Much might be said about such exotic principles. Note, for example, that *vv* is a plural variable. So we aren't reasoning from *sentences* in one language we speak to another, but supposing that we can (so to speak) reach out and catch the reference of a free variable in some formula in one language, by a formula in another language. But I take the general picture of accepting such inferences forcing a charitable interpreter to interpret the quantifiers in your new language E as ranging over strictly more objects than they did in your original Q to be clear. And I won't object to any of these details here. My objections concern the next part of the story.

With this background in place, Studd then considers how we can charitably interpret speakers who accept something like the inference rules above but have subtly incoherent beliefs[24]:

$$things(vv) \Rightarrow thing(\{vv\})$$

$$things(vv), v \prec vv \Rightarrow v \in \{vv\}$$

[23]See page 235

[24]See pg 239

$$things(vv), v \in \{vv\} \Rightarrow v \prec vv$$

The above inference principles let you infer that, for any plurality of things vv, there's a set $\{vv\}$ whose elements are exactly the objects v in this plurality vv (written $v \prec vv$). Thus accepting it (together with normal plural comprehension principles saying that for any ϕ there's a plurality vv of the objects such that ϕv) lets you derive the existence of the Russell set and hence contradiction.

Studd argues that these speakers could undergo a kind of unwitting quantifier meaning change, for the following reason. In general, a charitable interpreter can try to accommodate a speaker's reasoning by changing the domain of objects they take the speaker to quantify over[25] and the language they take them to be speaking. In this case, Studd suggests, charitable interpretation might take the speaker to be going through something analogous to the language switch from Q to E envisaged above. And if meaning reflects charitable interpretation, then we can have a kind of unwitting quantifier meaning expansion in this way.

This is, I take it, Studd's proposal for how it could be true that (unbeknownst to us) our current quantifiers range over some steadily growing range of sets. He puts it forwards as the "basis for an idealized account of universe expansion applicable to the ordinary English speaker". I have the following concerns.

[25] Studd gives this example, "I utter '52% of people voted for Brexit' and we immediately limit the domain to exclude those who didn't turn out or were ineligible to vote"

First, surely we don't actually, after the discovery of Russell's paradox, have the disposition to infer that arbitrary pluralities form a set.

Second, and perhaps most importantly as regards answering the double duty problem, Studd's story doesn't suggest any principled answer to when and how quickly speakers are supposed to go through language change events he proposes. How often would the charitable interpreter would say that someone with the inference dispositions above has switched languages (every 5 minutes? every 10 minutes?). If I lie around, having the inconsistent inference dispositions Studd mentions and not thinking about set theory for an hour, how many times should the charitable interpreter take my language to have changed during that time? Insofar as standing dispositions to make inferences (or regard failure to make inferences as irrational) drive the above charitable interpretation, it is hard to see how one could give any non-arbitrary answer to the above question.

Third, it's not clear whether *the balance* of charitable interpretation favors Studd's strategy, once we fill in relevant speakers' *other* inference dispositions in a realistic way. For one thing, people are disposed to interpret things they wrote yesterday homophonically, and assume that the truth value of sentences depending on the height of the sets doesn't change from day to date. But Studd's favored charitable interpretation would make this trans-language inference schema fail.

For another thing, it seems to me there's a dilemma about what different stages of growth in the hierarchy of sets are supposed to look like. If the hierarchy of sets grows one layer at a time, then it looks like reinterpreting someone as talking about a larger hierarchy will sometimes be very uncharitable. For example, doesn't going from interpreting someone as talking about V_ω to $V_{\omega+1}$ make various things they believe like the pairing axiom come out wrong?

But if we avoid this problem by saying that each reinterpretation of set theoretic talk must interpret people as quantifying over a domain of objects satisfying something like ZFC_2, we will 'ascend in big leaps' like Hellman rather than in single steps as Linnebo and I prefer, and face the inconveniences discussed in Chapter 3.

Also, at the risk of sounding crude, why isn't Putnamian potentialism (which, as we saw, Studd acknowledges the acceptability of) a more charitable interpretation than any of these? Why isn't the Parsonian interpretation itself a better interpretation?

So, overall I don't how to get any clear attractive answer to the question 'how many actualist sets am I currently thinking in terms of?' from Studd's account.

5.7 Conclusion

In this chapter I've discussed the differences between Putnamian and Parsonian approaches to Potentialism, and reviewed some major forms of Parsonian potentialism.

I've then tried to justify my use of a Putnamian framework to Parsonian readers who may find it unfamiliar. I've argued that Parsonians can and should accept the meaningfulness of basic concepts like logical possibility and can likely use my Putnamian version of replacement to further justify their version of replacement (at least to some extent). I've also (loosely) indicated the reasons why working in the Putnamian system will be practically useful.

More tentatively, I've argued that we should favor Putnamian over Parsonian approaches to potentialist set theory on approximately the following grounds.

First, Putnamian potentialism can be developed using a joint carving notion of logical possibility which everyone has reason to accept (and Putnamians have extra reason to accept). Thus (interpretationalist) Parsonian set theory can seem unparsimonious and needlessly revisionary, insofar as it requires adding an interpretational possibility operator to our ideology and make certain otherwise unneeded revisions to our philosophy of language (denying that 'precise intensions always determine precise extensions').

Second, the notion of interpretational possibility can seem like an unattractive choice of theoretical primitive. For facts about intepretational possibility generally would seem to be frequently indeterminate, highly disputed and/or unprincipled facts. And Linnebo and Studd's paritcular versions of this concept can seem ad hoc restricted to allow assumptions needed to justify set theory, . Thus one might favor Parsonian potentialism on grounds of conceptual parsimony. There is also a worry that the Putnamian potentialist needs to - in effect- invoke a Putnamian approach to the ordinals (a notion of arbitrary sequences reconceptualization events satisfying the axioms for being a well ordering), in which case adding a philosophically different Parsonian approach to the sets seems particularly unmotivated.

Third Parsonians faced a 'double duty' problem which Parsonian views avoid. For example, interpretationalist Parsonians face awkwardness about how to answer the question 'how many sets are you currently thinking in terms of?' that is not answered by Studd's picture of unintentional language change.

In closing I will mention three, weaker, motivations for favoring the Putnamian approach .

First, Putnamian paraphrases promise to make the intuitively close relationship between math and logic explicit (specifically the notion of logical possibility interdefinable with entailment).

Second, the *practical convenience* of working in a Putnamian framework (and cashing out set theory in terms of logical possibility) discussed in §5.5.2, might be taken as evidence for the philosophical correctness of this approach. Going Putnamian promises to lets us rationally reconstruct the justification for our set theoretic beliefs from premises that seem clearly true more directly, using fewer primitives.

Third, (although I personally think we should accept a broadly neo-Carnapian philosophy of language), it's worth noting that Putnamian potentialist set theory doesn't require us to accept this controversial philosophical thesis while (interpretationalist) Parsonianism does[26]. Thus philosophers who re-

[26]The interpretational Parsonian can't mean interpretional possibility in the familiar Tarskian sense where all interpretations choose their domains from among some fixed universe of objects, otherwise we will have a maximum size which all interpretations of the sets have to be found within. On such a view actualists apparent commitment to

ject such neo-Carnapian philosophy of language will certainly favor the Putnamian approach over the interpretationalist Parsonian one. And perhaps the same goes for philosophers who would prefer to leave few ‘hostages to fortune’ and avoiding entangling the philosophy of set theory with unrelated philosophical controversies.

---

an arbitrary stopping points, which potentialism promised to let us avoid, seems to get dragged back in.

Part II

In Part I we discussed some existing potentialist and actualist approaches to the foundations of set theory, and noted some problems for them. We saw that actualists face an arbitrariness problem related to the Buralli-Forti paradox. Potentialists face a problem about clarifying the intended meaning their possibility and extendability claims (e.g., the intended interpretation of the $\Box$ and $\Diamond$) in a way that supports their project. And both actualists and potentialists face a problem about justifying the ZFC axioms, especially the Axiom of Replacement. With this picture in place, we can now begin this book's positive project. We also introduced my preferred notion of conditional logical possibility and saw saw how if can both express standard claims about logical possibility and do the work of second order logic.

In Part II of this book I will develop my particular version of (Putnamian) Potentialist set theory using conditional logical possibility and argue that it lets us avoid many of the problems discussed above. As we have seen, potentialist paraphrases of set theory make claims about how it would be (in some sense) possible to extend an initial segment of the hierarchy of sets.

In chapter 6 I will give an informal summary of how conditional logical possibility (and first-order logic) lets us formulate a version of Putnam's potentialist set theory which differs from, and simplifies, Hellman's formulation in a few key ways. Specifically, I'll make it clear how conditional logical possibility allows us to completely eliminate quantifying-in to the logical possibility operator.

The remainder of this part will be devoted to providing a set of axioms for conditional logical possibility and arguing for the truth of these axioms.

Chapter 6

Purified Potentialist Set Theory: An Informal Sketch

In this chapter I will informally present my preferred version of potentialist set theory (using the notion of conditional logical possibility), and clarify some philosophical issues about it.

I will employ a version of Puntam's approach, but appeal to logical possibility specifically (much as Hellman does) rather than metaphysical possibility. So when I say that it would be possible to have an initial segment V, I will mean (something like) that it would be logically possible for the objects satisfy 'is a penciled point', 'is connected by an arrow' to form an intended-width initial segment of sets when considered under these relations[1]. By

[1] So, for example, although I may casually talk about the possible existence of initial segment structures V_i, I don't mean to assert that there are (or could be) special objects called structures, as e.g. Shapiro does. Or at least, I don't want to say that we need such objects to understand set theory. All talk about 'the possibility of a structure existing',

using non-mathematical relations we avoid having to presume there is an antecedently meaningful notion of set or other mathematical relation.

However there are a number of details to note.

6.1 Two Sorted Initial Segment Structures

First, I will only require my iterative hierarchies to satisfy IHW, not ZFC_2 as Hellman does, for the reasons discussed in §**??**. Doing this makes it convenient to admit the levels in our hierarchies of sets as primitive objects in their own right rather than rely on the (non-obvious) fact that Von Neumann ordinals can serve that function inside the sets. So my iterative hierarchies will have two kinds of (first order) objects playing two different roles: those of sets and ordinal levels. with sets being related to one another by elementhood, ordinal levels being related to one another by less than, and every set being 'available at' some ordinal level. To reiterate, on my current way of talking the ordinals are not themselves sets[2].

Thus, I will employ five relations (of any kind) to characterize the notion of initial segment: two one place relations playing the role of set() and ord(), and three two place relations playing the roles of $\in$, $<$ (ordinal ordering) and @ ('is available at', where a set x is available at a stage s if it has been constructed at or before stage s).

---

in the potentialist paraphrase strategy above is merely shorthand for claims about the possibility of there being objects which instantiate specific non-mathematical first order predicates and relations in a certain way.

[2] Of course, should one desire, one can prove that my ordinals can be uniquely identified with the sets forming the Von Neumann ordinals in my system.

So, for example, we might use the following first order properties and relations: ... is a penciled point, ...is a penciled star, ...is connected to.. by a dotted/dashed/solid arrow.

And I will define a formula $\mathcal{V}(\mathrm{set}, \mathrm{ord}, <, \in, @)$ which asserts the relations $\mathrm{set}, \mathrm{ord}, <, \in, @$ apply to a objects in such a way as to satisfy our conception IHW of an initial segment of sets (for brevity I will often simply call these initial segments).

6.2 Structure Preserving Not Obeject Preserving Extendability

Second I will use the notion of conditional logical possibility to talk about how one such hierarchy of sets like structure could extend another.

I will define a formula $V' \geq V$ which says that one such initial segment extends another (in the intuitive sense where one initial segment of the sets can extend another), where V abbreviates a list of relations $\mathrm{set}, \mathrm{ord}, <, \in, @$ and V' abbreviates $\mathrm{set}', \mathrm{ord}', <', \in'$ @ [3].

Now we can say that it's logically possible for an initial segment V to extended by an initial segment V' by simple holding fixed (the relations in) V. I adopt the following abbreviation for this frequently used expression.

$$\Diamond_V(V' \geq V) \overset{\mathrm{def}}{\leftrightarrow} \Diamond_{\mathrm{set},\mathrm{ord},\in,<,@}(V' \geq V)$$

[3] So $V' \geq V$ says that the objects satisfying $\mathrm{set}', \mathrm{ord}', <', \in'$ @ form an intial segment extending the initial segment formed by the objects satisfying $\mathrm{set}, \mathrm{ord}, <, \in, @$.

This might be read as saying[4] that it's logically possible (holding fixed the structure of V) for another initial segment V' to extend V.

6.3 Assignment Functions and Hierarchies

Third, to completely eliminate quantifying in I won't just think about how an initial segments can be extended, but rather how initial segments augmented with a 'function' representing the assignment of variables can be extended. Thus, rather than talking about what's possible given an initial segment V and the object bound to the variable x as Hellman would, I talk about what's possible given an initial segment V and an 'assignment function' ρ, where $\rho(\ulcorner x \urcorner)$ is meant to capture the assignment of the variable letter 'x'.

In particular, I'll associate each initial segment V with a copy of the natural numbers $\mathbb{N}$ and assignment function ρ assigning 'numbers' to sets in that initial segment. Call the resulting structure an Interpreted initial segment. And let $\vec{\mathscr{V}}\left(\vec{V}\right)$ abbreviates the conjunction of the requirement that V is initial segment, $\mathbb{N}$ is a copy of the natural numbers and ρ is a function from $\mathbb{N}$ to the sets in V[5].

So, recall that to give a Hellman style potentialist translation of a sentence like $(\forall x)(\exists y)\phi(x, y)$ where ϕ is quantifier free we want to say something like

[4] In even more detail, it might be read as saying: It's logically possible (given the structure of the pencil points and arrows etc.) the pen points and arrows etc. form an initial segment extending an initial segment structure formed by the pencil points and arrows.

[5] Here I use variables $\vec{V}, \vec{V}'$ to abbreviate corresponding lists of relations $V, \mathbb{N}, \rho$ and $V', \mathbb{N}', \rho'$

this.

It's logically necessary that however one chooses a set x from an initial segment V, it's logically possible to extend this V with an initial segment V' containing a set x making $\phi(x, y)$ true.

We can capture the same content as the above sentence as follows[6].

$$\Box\left[\mathscr{V}\left(\vec{V}\right) \rightarrow \Diamond_{\vec{V}}[\vec{V}' \geq_y \vec{V} \wedge \phi(\rho(\ulcorner x \urcorner), \rho'(\ulcorner y \urcorner))][\in / \in']\right]$$

Here $\ulcorner x \urcorner, \ulcorner y \urcorner$ are objects in $\mathbb{N}$ [7] coding the variable 'x' and 'y'. And $\vec{V}'$ **extends another ensemble $\vec{V}$except on 'y'** (written $\vec{V}' \geq_y \vec{V}$) says that the initial segment $V' \geq V$ in the usual way, (relevant copies of the numbers $\mathbb{N}$ and $\mathbb{N}'$ are coexitensive[8] and) the assignment function ρ' maps $\mathbb{N}'$ to the sets in V' such that $\rho(n) = \rho'(n)$ for all numbers n *except* $\ulcorner y \urcorner$.

Ignoring the details for the moment, the key insight here the initial logical necessity operator lets ρ range over all possible relations, so the consequent must hold given any possible set (position) in V chosen by $\rho(\ulcorner x \urcorner)$.

So the claim above says (in effect) any way that $\rho(\ulcorner x \urcorner)$ could choose an 'x' in an initial segment V, it would be logically possible freezing this choice, to have an extending interpreted initial segment $\vec{V}'$ with a ρ' assigning 'x' the same way and 'y' so that $\phi(x, y)$ holds between the objects assigned to 'x' and 'y' respectively.

[6]Note, here I use functional notation for ρ i.e., I write $\rho(x) = y$ rather than $\rho(x, y)$

[7]$\ulcorner x \urcorner$ is represented as $S(S(S(\ldots S(0)))$ for some number of successor operators and 0 s the unique element of $\mathbb{N}$ that isn't a successor and S is a relation that we write functionally.

[8]That is, all relations in $\mathbb{N}'$ are coextensive with corresponding relations in $\mathbb{N}'$

6.4 Atomic Predicate Use Reducing Trick

One final question which naturally arises is whether this style of potentialist paraphrase requires appeal to infinitely many atomic predicates. As stated so far, my strategy would require access to infinitely many atomic relations if we want to be able to translate set theoretic sentences with arbitrarily deep nested quantifiers. For instance, the potentialist translation of $\forall x \exists y \forall z \phi(x, y, z)$ would seemingly require three distinct tuples of relations $\vec{V}, \vec{V'}, \vec{V''}$.

However, a careful examination of our translations shows that we only preserve the relations from the prior possibility context. Thus, if desired, in the above potentialist paraphrases we can replace V^n with $V^{n \mod 2}$ (where V^1 is just V', V^2 is V'' etc..) without affecting the truth value of the translation. This allows us to translate sentences with arbitrarily many quantifier alternations using a fixed finite number of atomic relations.

Here's what I mean. We translate a sentence with three quantifiers $\forall x \exists y \forall z \phi(x, y, z)$ as follows:

$$\Box(\vec{\mathscr{V}}(V) \rightarrow \Diamond_{\vec{V}}[\vec{V'} \geq_y \vec{V} \land \Box_{\vec{V'}}(\vec{V''} \geq_z \vec{V'} \rightarrow \phi(x, y, z))]$$

But note that logical possibility treats all relations of the same arity the same. And conditional logical possibility treats all relations (that aren't being held fixed) of the same arity the same. So this assertion :

$$\Box_{\vec{V}'}(\vec{V}'' \geq_z \vec{V}' \to \phi(x, y, z))$$

is true if and only if:

$$\Box_{\vec{V}'}(\vec{V} \geq_z \vec{V}' \to \phi(x, y, z))$$

That is, replacing V'' with V has no effect. So we can formalize the same claim like this.

$$\Box(\vec{\mathscr{V}}(V) \to \Diamond_{\vec{V}}[\vec{V}' \geq_y \vec{V} \wedge \Box_{\vec{V}'}(\vec{V} \geq_z \vec{V}' \to \phi(x, y, z))]$$

For readability I'll write as if I have access to an infinite number of distinct relations of each arity. But keep in mind that the argument above demonstrates we can limit ourselves to only 16 distinct atomic relations.