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 as 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 that 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] Talk about the height of such a hierarchy of sets refers to the "number" of layers, while talk about its width refers to how many sets are introduced at each stage.

One can spell out this idea of a full-width iterative hierarchy as follows.

[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 Section 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.

https://doi.org/10.1017/9781108992756.002 Published online by Cambridge University Press

Definition 2.1 (Iterative Hierarchy – Full Width (IHW)). A full-width iterative hierarchy (IHW) is a structure consisting of:

- a well-ordered series of levels; and
- a collection of sets "available at" these levels, such that:
  - at each level, 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 IHW 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. One can (clearly) formalize the above claim using second-order logic, and I'll refer to the resulting theory as IHW_2.[2]

In contrast the ideas evoked above do not pick out a unique intended height for the hierarchy of sets.[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 the version of Potentialism I favor will wind up denying that there is, strictly speaking, a hierarchy of sets (hence anything for mathematical talk of "the height of the hierarchy of sets" to refer to[4]).

2.2 A Burali-Forti Problem

The problem for actualist set theory is not simply that it might be impossible to define the notion of absolute infinity in other terms. After all, every theory will have to take some notions as primitive.

Instead, we find ourselves in the following situation:

- Our naive conception of absolute infinity (the height of the Actualist hierarchy of sets) turns out to be incoherent, not just unanalyzable.
- And, once we reject this naive conception, there's no obvious fallback conception that *even appears* to specify a unique height for the hierarchy of sets in a principled and sufficiently clearly consistent 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

[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 3. I'll understand IHW loosely to be compatible both with a Boolos style two-sorted conception and the standard cumulative hierarchy.

[3] We could add the principle that there is no last stage, as Boolos (1971) 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.

https://doi.org/10.1017/9781108992756.002 Published online by Cambridge University Press

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 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 for every set/ordinal *x* there's a strictly larger set/ordinal *y*. Nor do I mean to suggest that there could somehow be "sets beyond all the sets," or that there's something wrong with taking various concepts used in articulating a conception of the hierarchy of sets as primitive (it's hard to see how one could avoid doing the latter!).

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. But 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 or combinatorically if not metaphysically) possible for there to be a strictly lager model of IHW which, in effect, adds a new stage above all the ordinals within the original structure together with a corresponding layer of classes.[5] And, worryingly, 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 structure. The hierarchy of sets just happens to have some particular height, although nothing in our conception of the sets rules out epistemic possibilities where the hierarchy of sets is taller.

But saying that the hierarchy of sets just happens to stop at a certain point seems to violate intuitive principles of metaphysical parsimony. It seems to require acknowledging 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: why should we think set theorists' reasoning about large cardinals etc., correctly reflects this brute fact about where the hierarchy of sets happens to stop?

[5] I won't say more about how to spell out the informal notion of possibility being invoked here now, but each version of Potentialist set theory discussed below (mine included) brings with it a candidate modal notion.

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).[6] However, there's no obvious fallback/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 Shapiro and Wright (2006) put it, 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. So, the completeness theorem doesn't guarantee that our conceptions of these structures will have "intended" models in the hierarchy of sets (i.e., 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 we 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.[7]

Note that, if some Actualist claimed to have a suitably primitive and seemingly precise notion of absolute infinity, they wouldn't face the arbitrariness worry I'm pressing. They could appeal to this notion of absolute infinity to specify the height of the hierarchy of sets. 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. They don't claim to have a concept that seems capable of picking out a precise intended height without deference to prior facts about however tall a hierarchy of sets there happens to actually be. Arbitrariness troubles arise because we start out with the seemingly precise naive conception of the intended height of the hierarchy of sets, and no other seemingly precise notion appears to fill the gap once this naive conception is rejected as paradoxical.[8]

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

[7] See Hellman (1994) for a version of this generality worry.

[8] That is, I take it most Actualists would agree that we don't even *seem* to have an independent precise (primitive or otherwise) conception of the intended height of the hierarchy of sets in the way that (many

https://doi.org/10.1017/9781108992756.002 Published online by Cambridge University Press

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, it might seem natural to say that the sets are the shortest possible structure satisfying ZFC_2 (i.e., the hierarchy of sets, so to speak, stops below the first inaccessible). This proposal is natural as it mirrors how we pick out a unique structure for the natural numbers by saying that the numbers are "as short as can be" while being closed under successor. 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.

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," McGee (1997) advocates a conception of an iterative hierarchy of sets with ur-elements, and proves a "quasi-categoricity" theorem about it, which might seem to help the Actualist address the arbitrariness challenge posed above.

However, I will argue that this is an illusion. Although McGee's characterization of a hierarchy of sets 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.

McGee (1997) 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).

would say) we do *seem* to have a conception of the intended width of the hierarchy of the sets or what second-order collections or pluralities there are supposed to be. An Actualist who (unlike all the Actualists I've encountered) did claim to grasp a primitive notion of absolute infinity that picked out a precise structure in this way would not face the arbitrariness problem above. See Section 2.5.2 for much more detail regarding this distinction.

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 schemas 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. One keeps the core idea of an iterative hierarchy of sets described above (with each layer containing "all possible subsets" from the lower layers), but takes 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 than just the empty set. 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 Ur-element Set Axiom follows from the statement above, and 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 $\text{ZFC}_2 + U$ (the result of adding the above ur-element principle to second-order ZFC set theory) has a property which he calls "quasi-categoricity."[9] 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 simulaneously: choose "sets" from within this domain, take quantifiers to range over this whole domain and make McGee's $\text{ZFC}_2 + U$ 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 $\text{ZFC}_2 + U$ 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's 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 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

[9] One might worry about the above axiom on the basis of Uzquiano's proof that McGee's axioms for set theory with ur-elements are incompatible with certain axioms of mereology (Uzquiano 1996), but I leave this question aside as the concerns I will be raising are unrelated.

https://doi.org/10.1017/9781108992756.002 Published online by Cambridge University Press

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 even show that $ZFC_2 + U$ is *quasi*-categorical in the following (to my mind, more natural) sense of the term. It doesn't show that, fixing the facts about what non-set objects there are, any hierarchy of sets satisfying $ZFC_2 + U$ must have a certain unique structure. Indeed, given certain popular assumptions you can always[10] 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.

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 Martin (2001) might at first sound like it helps the Actualist with the arbitrariness/lack of a definite conception worry, but actually does not. Indeed, Martin seems to positively endorse a version of this worry (Martin n.d.).[11]

[10] For instance, if we presume the existence of unboundedly many inaccessibles, as is often thought plausible, we are guaranteed multiple models of $ZFC_2 + U$ with a particular collection of ur-elements.

[11] There 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 *x* s;
3. the concept of transfinite iteration;
4. the concept of absolute infinity.

Perhaps we should include the concept of Extensionality as Component (0).

https://doi.org/10.1017/9781108992756.002 Published online by Cambridge University Press

Martin (2001) argues against plenitudinous anti-objectivist “multiverse” approaches to set theory (like Hamkins 2012) 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 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 *V*s 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 by noting that if we accept a certain conception of the hierarchy of sets (including the principles below), we can derive that there could not be two different hierarchies of sets (“the red sets” and “the blue sets”):

- 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 points out 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 ur-elements, and then on their first layer and their second layer etc.

One can call this a categoricity result. But it doesn’t answer our worry about arbitrary stopping points. For it doesn’t imply that it’s logically or metaphysically necessary that any collections of objects which satisfy the above 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

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. (Martin n.d.)

https://doi.org/10.1017/9781108992756.002 Published online by Cambridge University Press

satisfying ϕ_κ. It merely shows that we couldn't have two actual hierarchies of sets satisfying Martin's assumptions, 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 the hierarchy of sets has, allows it to satisfy 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 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:

$$\forall w_1 \forall w_2 \ldots \forall w_n (\forall a [\forall x (x \in a \rightarrow (\exists! y) \phi(x, y, w_1, \ldots, w_n))])$$
$$\leftrightarrow \exists b \forall x ((x \in a \rightarrow \exists y (y \in b \wedge \phi(x, y, w_1, \ldots, w_n))))]$$

So, Replacement says that 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 (1971) points out, 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 assumptions in IHW plus the extra claim that 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 with $\omega + 1$, 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. As mentioned in the introduction, Hilary Putnam (2000) 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."

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,"[12] and then reports of our present situation that, "In the case of Replacement there is, it is true, no 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'" (Potter 2004).

To my knowledge, four main (Actualist) strategies for justifying Replacement are currently popular. First, one can try to justify the axiom of Replacement "extrinsically" in the way we often 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. See Koellner (2009) on Gödel's proposal:

> 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.

However, 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 everything we want to say about the natural numbers (in the language of arithmetic) follows from (say) our second-order conception of the natural numbers.

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 perhaps 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 ZF axioms, which are treated as quite secure and used to provide a foundation/explication of normal mathematical claims that we are very confident in. Even if appeal to the fruitful good consequences of Replacement provides some justification for believing it, this doesn't secure the kind of intrinsic convincingness we usually expect (and hope for) from mathematical axioms.

Second, Potter (2004) suggested justifying Replacement by appeal to a kind of inference to the best explanation along the following lines. Russell's paradox tells us

[12] Potter 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 $\{\omega, P(\omega), P(P(\omega)), \ldots\}$ cannot be proved to exist in that system; yet 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'."

https://doi.org/10.1017/9781108992756.002 Published online by Cambridge University Press

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 which can form 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 (Potter 2004). As Michael Potter puts it, we should accept the 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 Replacement), 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? (Potter 2004: 231)

I don't find this inference to the best explanation very convincing because sets do have *some* other features than their size which could be used to explain why certain pluralities of sets fail to form a set in a style analogous to Potter's explanation for this fact.

In particular, note that on 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 is 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)[13] 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 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, if we're just accepting Replacement on the basis of inference to the best explanation, 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 the

[13] See, for example, Bliss and Trogdon (2016) for the development of the intuition that the existence of Socrates's 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.

https://doi.org/10.1017/9781108992756.002 Published online by Cambridge University Press

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 in typical mathematical results.

Third, one can provide a kind of justification for Replacement by noting it follows from a set-theoretic reflection principle.[14] 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) rather than any 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 true in some proper initial segment 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.[15]

This third strategy (justification by appeal to a reflection principle) is *somewhat* attractive. For, as Koellner (2009) reviews, one can motivate reflection principles[16] 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 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[17] 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.[18]

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).

[14] My summary of this approach follows Koellner (2009).

[15] See, for example, Button (n.d.).

[16] 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.

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

[18] This is quoted from Wang (1998) in Koellner (2009).

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, IHW).

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 (Reinhardt 1974). So, one might think that justifying Replacement by merely noting that it follows from Reflection doesn't provide enough justification.

Fourth, philosophers like Boolos (1971a, 1989) 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.

A fifth style of justification considered by Button (n.d.) derives Replacement from the following principle:

Stages-are-super-cofinal. If A is a set and $\tau(x)$ is a stage for every $x \in A$, then there is a stage which comes after each $\tau(x)$ for $x \in A$.

Button notes that we can motivate the following formal claim by appealing to the informal principle below, which he says is "consonant with" the cumulative-iterative conception of set:

Stages-are-inexhaustible. There are absolutely infinitely many stages; the hierarchy is as tall as it could possibly be.

However, I don't currently grasp the kind of modality that's intended to be evoked by the term "possibly" in *stages-are-inexhaustible*. Earlier in this chapter I've tried to invoke an intuitive sense of possibility on which there *couldn't* be an iterative hierarchy "as tall as it could possibly be" (for any structure of objects satisfying IHW, there could be a strictly taller one). And we will see below that Potentialists like Putnam, Parsons, Hellman, Linnebo and Studd have appealed to notions of logical or interpretational possibility which (they think) conform to this intuition.

And without the additional justification provided by the informal principle *stages-are-inexhaustible*, we find ourselves in an epistemic situation similar to just taking Replacement or some form of Reflection as an axiom, as regards *stages-are-super-cofinal*. It's not implausible, but also not seemingly obvious/clearly true that it would be logically coherent for there to be an iterative hierarchy that satisfies the relevant closure principle. Thus, I don't think merely pointing out that *stages-are-super-cofinal* implies Replacement doesn't suffice to justify the latter from principles that seem clearly true.

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 real numbers 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, we seem to have a unified, precise and intuitively consistent conception of the relevant mathematical structure, from which our first-order axioms describing the natural numbers/real numbers flow.

The iterative hierarchy idea sketched in Section 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 they satisfy Replacement). However, we have little or no reason to think this enlarged conception is coherent. 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 adopting 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 by quickly saying something about the limits of the argument discussed in this chapter.

Many other philosophers interested in Potentialism about the height of the hierarchy of sets, such as I will develop in response to the arbitrariness worry, 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 Potentialist approach to the height of the hierarchy of sets but not to the width of the hierarchy of sets or the natural numbers.

[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.

https://doi.org/10.1017/9781108992756.002 Published online by Cambridge University Press

In the remainder of this chapter, I will answer this question by clarifying why I think the motivation for height Potentialism about set theory doesn't generalize in the ways just mentioned. 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

While one can certainly doubt that we can uniquely refer to the intended structure of the natural numbers or the subsets of a given set there is no (similarly compelling) paradox like Burali-Forte that arises from assuming that we can.

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 satisfying IHW. We can specify exactly how $\leq$ and $\in$ would relate the new sets/ordinals to all the old sets/ordinals previously considered so as to form a new structure satisfying IHW equally well. And the structure we imagine forming by extending any given plurality of ordinals has as good a claim to contain all the objects that satisfy our conception of "the ordinals"/"the sets" as the original structure, if our conception after rejecting the naive height principle is just IHW. And in any case 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.[21]

[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.

[21] Perhaps Hamkins' radical multiverse proposal provides a way of developing the latter counter-intuitive idea. But see my discussion of Hamkins in Section 9.4.

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 "What Is Mathematics About?," Dummett (1993) raises something very much like the Burali-Forti worry I pressed above concerning the height of the hierarchy of sets:

> If it was . . . all right to ask, "How many numbers are there?", in 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.[22]

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) Platonism about set theory because our concept of sets and ordinals is "indefinitely extensibile" in the following strong sense (if we take the natural conception of set that remains, once we reject the naive and paradoxical conception that the sets go "all the way up," to be IHW):

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 conception of the sets (IHW)/ordinals equally well.

In contrast, Dummett seems to reject standard Platonist set theory because our concept of sets is "indefinite extensible" in this weaker sense:

Weak Indefinite Extensibility For any collection of numbers/sets/ordinals *we can form a definite conception of* (which Dummett 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

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" (Dummett 1993: 440; emphasis added).

To support this reading, consider how 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

[22] Dummett (1993: 439).

https://doi.org/10.1017/9781108992756.002 Published online by Cambridge University Press

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:

> 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 1993: 443)

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.

I won't try to adjudicate this dispute here. Much can and has been said about whether this succeeds and how to understand Dummett's infamously "dark" (Rumfitt 2015) notion of indefinite extensibility (Dummett 1991).

Instead, I merely want to note that Dummett's arguments for the (weak) indefinite extensibility of the natural numbers and real numbers don't even pretend to show the strong indefinite extensibility 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 it would be intuitively possible/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 relating your natural number concept to reference magnetic notions of second-order quantification or logical possibility.[23]

Thus, I think, the fact that Dummett's more skeptical worry applies more widely than the Burali-Forti driven worry I've pressed is unsurprising.

[23] Perhaps 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, Axiom 8.1, and uniform relabeling, Axiom 8.5, principles, I introduce below in 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, or reduces to my worry about the height of the hierarchy of sets. 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.

https://doi.org/10.1017/9781108992756.002 Published online by Cambridge University Press