Conceptualization as Explanatory Coherence: an Application to Set Theory's Potentialism vs. Minimalism Debate

MSc Thesis *(Afstudeerscriptie)*

written by

Jori Koolstra
(born November 3, 1995 in Groningen, the Netherlands)

under the supervision of **Dr. Thomas Schindler**, and submitted to the Examinations Board in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
April 26, 2024	Prof. Maria Aloni (Chair)
	Prof. Luca Incurvati
	Dr. Giorgio Sbardolini
	Dr. Thomas Schindler



Institute for Logic, Language and Computation

Abstract

This thesis is concerned with the philosophy of set theory, and therefore with the foundations of mathematics. Our standard theory of set, **ZFC**, is rooted in the iterative conception, according to which sets arise as collections *of some things*. Boolos [12] (1971) inaugurated the modern study of conceptions of set, emphasizing their justificatory role in deriving axioms and avoiding the paradoxes of naive set theory. More recently, Incurvati [54] (2020) conducted an extensive comparative analysis of several conceptions of set, advocating for the correctness of the iterative conception based on certain virtues it possesses over the other conceptions discussed.

Different accounts of the iterative conception have been proposed in the literature. Of particular interest are Linnebo's 2010 and 2013 papers [69] [68], which introduce a potentialist perspective on iteration formulated in modal plural logic. According to Linnebo, the conventional language employed in set theory lacks the necessary granularity to properly address certain foundational concerns, such as the paradoxes. This view has, however, recently been challenged by Soysal [100] (2020). She argues that e.g. V is not a set simply because its existence conflicts with the axioms supported by the iterative conception (and this conception is correct). If this conception-based explanation goes through, Linnebo's modal apparatus may turn out to be an 'idle wheel', as Soysal's argument applies even to the minimalist account of the iterative conception, according to which its content is exhausted by saying that the sets are those objects that occur in the cumulative hierarchy. Therefore, to evaluate Linnebo's conception of set, we need a general theory of conceptualization in mathematics on which we can compare such different accounts.

I introduce a Lakatosian account of conceptualization, drawn from Incurvati [54], on which conceptions arise when open-texture in a *foundational* concept is sharpened. Due to this foundationality, we must exercise careful control on stipulation in conception formation, in particular to handle circular justification, for instance regarding the height of the cumulative hierarchy. The novel solution I propose is to derive conception-admissibility from *explanatory coherence*. This can be achieved by answering as many (why-)questions as possible. Moreover, I will argue that *explanatory richness* in this sense is also a virtue that should weigh on a conception's correctness. Finally, I show that Soysal's minimalist answer to why V is not a set is explanatorily unstable, unlike the solution derived from Linnebo's potentialism. I will therefore conclude that his modal apparatus does genuine explanatory work, and thus the potentialist account must be preferred as more correct.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International license.

Acknowledgments

Michael Sandel in *The Tyranny of Merit* has emphasized that our accomplishments do not occur in a vacuum, but are as much the product of the support we have enjoyed throughout our lives as of our own abilities and effort. I can only concur with this view. These acknowledgments are long overdue and I have so many people to thank.

I'm grateful to my supervisor, Dr. Thomas Schindler. Without our discussions I would not have been able to reach the clarity I needed to lay out my arguments. I am also indebted to my previous supervisor Dr. Giorgio Sbardolini, to my teacher Prof. Luca Incurvati at the ILLC, and to Prof. Michael Hallett, my favorite professor at McGill who first taught me philosophy of mathematics in Montreal, and who encouraged me to study logic. I did, and here is the result. Thank-you all for sparking my passion for philosophy.

Thank-you Tex, *for everything.* You're my favorite sparring partner and offline encyclopedia, accessible at 3pm and 3am, in CET and EST. You're an absolutely irreplaceable and dear friend. I really doubt this thesis would have been here without your support. There are few words to convey the depth of my gratitude, but perhaps a dinner at an Iraqi restaurant is a start…? And the same holds for you, Jochem. Maybe one day I will achieve true chimpdom and attain the ultimate $\{\bar{\sigma}\}$. If so, it will be because of your encouragement, humor, and lifelong friendship. I love you two very much!

Thank-you Jeanette and Tom for your hospitality and advice. I was always welcome when needed. Special thanks to Paul as well for all the skipskipskibbediedoos and the invaluable cooking tutorials. I know now to always double the garlic and never cook chicken!

I further wish to express my gratitude to the brothers Ahmed, Omar, and Mohammed Ellithi, and to Hany and Nasr, as well as to Fuat, Furqaan, and Rusydi, from who I have learned so much. I will see you in Indonesia soon! Thank-you Freek, Justin, and Elyas. I must also make special mention of Mahmood Mansour for his extraordinary kindness and hospitality. Thank-you my dear brother.

Special thanks to Hossein Abukar and Mohammad Awad. I cherish my memories of cruising through Atlanta with y'all and eating together. Hossein, your presence illuminates the lives of those fortunate enough to know you. Thank-you for being my friend and mentor when I was new to America and knew nobody. Mohammad, you're truly one of a kind. I've learned so much from you in such a brief period of time. You are an example for those around you. Thank-you for everything. The same goes for Mohammad from Casablanca. Although we only met a couple of times, the inviting example of openness and genuine kindness is

something that has stuck with me ever since we first met in Lilburn. I've never had a private tour of a water treatment plant! I hope to join you all there one day.

Sonam, you listened to me when I needed it the most. You understand what most people do not. Keep going, you got this. I am rooting for you! Thank-you for showing me the beauty of the Blue Ridge Mountains. Maybe one day we will still catch those stars? Georgia on my mind!

Thank-you everyone for making me feel at *home* wherever I was.

Finally, I couldn't have finished this thesis without Tanja's ongoing encouragement and support. Thank-you sincerely for your exceptional care and attention, which was well beyond what I could have ever reasonably expected. You really made all the difference.

This thesis is dedicated to you, Mom. Without your love, I wouldn't have even be able to start it.

Jori.

Groningen,
Amsterdam,
Lawrenceville, Georgia
(United States),
February 2024.

Contents

Introduction **1**
0.1 Where Set Theory is Today 1
0.2 The Iterative Conception 3
0.3 Looking Forward 4

1 Set Theory in Mathematical Practice **7**
1.1 Replacement 7
1.1.1 Extrinsic and Intrinsic Justification 7
1.1.2 Is Replacement Necessary? 9
1.1.3 The Heuristic Use of Replacement 10
1.1.4 The Foundational Role of Replacement 12
1.1.5 Conclusion 14
1.2 Naive Comprehension 14
1.2.1 Cantorian Set Theory and Paradox 14
1.2.2 Skeptical Remedies 18
1.2.3 Mathematical Richness 20
1.2.4 Conclusion 21

2 Conceptions and the Significance of Why-Questions **23**
2.1 Concepts and Conceptions 25
2.1.1 Open-Texture 27
2.1.2 Foundational and Instrumental Concepts 33
2.2 Conceptual Correctness 38
2.2.1 The Conception-Admissibility Problem 38
2.2.2 The Dilemma of Conceptual Correctness 40
2.3 Stipulation, Freedom, and Coherence 42
2.3.1 The Hilbertian Criterion 43
2.3.2 The “Use in Practice” Criterion 44
2.4 Explanatory Coherence and Why-Questions 46
2.4.1 The Role of Explanatory Saturation in Coherence Arguments 46
2.4.2 Explanatory Coherence in Mathematics 49
2.4.3 Why-Questions 51
2.4.4 The Justification of Infinity and Measurable Cardinals . . . 52

2.4.5 The Tractability of Conceptual Justification and its Demarcation from Extrinsic Evidence 53
2.5 Conclusion 55

3 Why is V not a Set? **57**
3.1 What Makes a Good Explanation of the Paradoxes? 58
3.2 Linnebo's Potentialist Account 60
3.2.1 The Generative Metaphor and Ontological Dependence . . . 61
3.2.2 The Potentialistic Conception 63
3.2.3 Modal Set Theory 66
3.2.4 Some Problems with Linnebo's Potentialism 70
3.3 Conclusion 72

4 Potentialism versus Minimalism **75**
4.1 Idle Wheel or Resolution Magnifier 75
4.2 Critique of Minimalism 77
4.3 Conception-Based Minimalism 80
4.4 Inference to the Best Conception 86
4.5 Soysal's Minimalism is Explanatorily Unstable 87
4.6 Conclusion 90

Introduction

In the next section we briefly discuss some relevant history of set theory. Then, in Section 0.2, we introduce the iterative conception of set and some important problems surround it that must be addressed before this conception can qualify as a satisfying account of set. We also indicate how these considerations connect to Linnebo's potentialist account. Finally, Section 0.3 gives an overview of the structure and content of this thesis.

0.1 Where Set Theory is Today

The theory of sets, understood as starting with Cantor (1845–1918), serves modern mathematics in multiple integral roles:[1]

- as a universal background language, to allow interchange and application between different mathematical fields;
- as an arbiter, an ultimate axiomatic foundation and judge of proof for classical mathematics;
- as a versatile scale of mathematical strength;
- as a theory of the higher infinite.

For the modern mathematician, the standard first-order axiomatization **ZFC** (*Zermelo-Frankel set theory* with the axiom of Choice) has distinctly consolidated these promises. Moreover, if not already considered as an outright true theory of set, **ZFC** is generally assumed to have a fairly wide 'safety margin'. It is supposed that even if some of the outer fringes of its infinitary implications turn out to be false, any resultant inconsistency can be resolved locally and will not seep through to the more concrete subsystems required for developing (classical) mathematics. In other words, the (implicit) general belief among mathematicians nowadays is

[1] Maddy [76] [77] identifies several foundational desiderata exemplified by set theory. In addition to the first three roles of set theory mentioned here, which correspond respectively to her "Generous Arena", "Shared Standard", and "Risk Assessment", she also suggests "Meta-mathematical Corral" and "Elucidation".

that outside of strictly set-theoretic woes, or concerns that are otherwise of meta-mathematical origin, **ZFC** as a foundation for mathematical practice is in excellent shape. There is no *mathematical* need for continued study and re-reassessment.

This situation was very different in the late 19th and early 20th century. As set-theoretic techniques were more and more being put to work in mathematical arguments, numerous paradoxes in the naive use of sets surfaced. This dissonance propelled a fierce *Grundlagenstreit* between various factions of friends and foes of the new set-theoretic methodology, wedged in particular on issues like impredicativity, the status of nonconstructive reasoning, and the use of completed infinities. Foes of set theory identified in these innovations the cause of the paradoxes, while friends pointed to their wonderful applications, such as in Cantor's proof of the existence of transcendental numbers and in the proof of the Cantor–Bendixson Theorem. In this effervescent context Hilbert (1862–1943) famously proclaimed that "No one shall drive us out of the paradise which Cantor has created for us."[2]

When the friends of set theory ultimately prevailed and **ZFC** was later consolidated as a powerful but apparently consistent theory, foundational issues were pushed to the back as a matter to be clarified by logicians and philosophers. The foundations of mathematics have remained under their guardianship ever since.[3] Colyvan puts it thus: "Philosophy of mathematics kept going [after the early 20th century], of course—philosophy always does—but it had lost its urgency and, to some extent, its raison d'être."[4] Still, regardless of any engagement with mathematical practice, there certainly is a strong *philosophical* interest to clarify and study the conceptual underpinnings of **ZFC** and related systems. If **ZFC** is supposed to be *about* sets, then as philosophers we ought to have an account of the concept of set, and that account ought to *explain* which sets there are and of which we shall not speak, how they come to us, and what principles to accept regarding them and why.

[2]Hilbert [48], p. 191.

[3]A notable recent exception to this is the Fields medalist Voevodsky, who developed *univalent foundations* in response to the increasing need of modern mathematics to handle "Proof Checking" at the foundational level (see, e.g., Maddy [77], pp. 304-10). Another example is the indefatigable ultrafinitist Edward Nelson, who produced several (flawed) proofs of the inconsistency of infinitary arithmetic (see his [82] for an introduction to his views).

[4]Colyvan [25], p. 3.

0.2 The Iterative Conception

Such conceptual underpinnings for **ZFC**-like theories are today predominantly provided by the *iterative conception of set*.[5] This conception of set "according to which a set is anything obtainable from [the empty set] by iterated application of the operation "set of," [...] has never led to any antinomy whatsoever."[6] Faith in this conception of set is further strengthened by the imagination of an intuitive model, namely the familiar picture of an upside-down pyramid of stacked layers of sets ordered by the subset relation, so that the empty set is at the bottom. The basic thought is that, unlike on the naive view, sets do not exist independently of each other as *logical* extensions of predicates, but are always *combinatorially* coalesced as *sets of some things*. For any collection of urelements U (possibly empty) one obtains their set-theoretic hierarchy by iterating the 'set of' operation on its earlier outputs, initiating with U. When iterated in this way, starting from the empty set, we 'generate' the familiar *cumulative hierarchy* of pure sets:

$$\begin{aligned} V_0 &= \emptyset\,; \\ V_{\alpha+1} &= \mathcal{P}(V_\alpha)\,; \\ V_\lambda &= \bigcup_{\alpha<\lambda} V_\alpha \text{ for limit } \lambda\,. \end{aligned}$$

Regardless of whether the sets are actually structured as in this hierarchy, the iterative conception gives rise to explanatory problems about its interpretation of the concept of set.

Problem 1. For the iterative conception to be a satisfying account of set, we need to clarify several foundational questions:

1. how far the cumulative hierarchy extends;
2. what principles it can justify;
3. what relationship exists between the (layers of) sets;
4. how the paradoxes of naive set theory are explained and defused.

This thesis is inspired by Linnebo's [68] [69] recent attempts to show that set-theoretic potentialism—the idea that sets exist only potentially in relation to

[5] See in particular Boolos's [12] and [13]. The distinction between concept and conception is laid out in Section 2.1.

[6] Gödel [39], pp 518-9.

their members—can help shine a light on some of these questions. However, before we delve into the main themes of this thesis, it is worth to consider the broader relevance of the upcoming discussion. I will argue that, against the common belief among mathematicians, the foundational enterprise remains intertwined with mainstream mathematics. The first goal, then, is to relate the following two set-theoretic principles to mathematical practice. These principles are of particular relevance to Linnebo's "finer resolution" approach to set theory (Section 4.1), as they involve especially divisive aspects of the conceptualization of set.[7] Their status also pertains to the questions posed above, as they prompt considerations about...

1. whether the cumulative hierarchy extends far enough to support **Replacement**,

$$F \text{ is a function} \to \forall z \exists y \forall x (x \in y \leftrightarrow \exists w (w \in z \wedge F(w) = x)) .$$

   (this ties into questions 1 and 2);

2. whether the iterative conception can supply an explanatory account of the paradoxes associated with **Unrestricted Comprehension**,

$$\exists x \forall y (y \in x \leftrightarrow \varphi(y)) ,$$

   that fares better than other conceptions of set with respect to certain explanatory qualities. (this ties into questions 3 and 4.)[8]

Linnebo's potentialistic view of sets in fact emerged from his diagnosis of the set-theoretic paradoxes as a consequence of the inadvertent collapse of the finer modal underpinnings of set theory to fit the common actualist narrative that there are all the sets. According to Linnebo, reintroducing this modality allows Unrestricted Comprehension to be restored to its true non-pathological form. The modal language furthermore suggests new "natural principles," such as modal reflection (Formula $\Diamond$-Refl), which provides a novel avenue to argue for Replacement.

0.3 Looking Forward

In Chapter 1 we identify and discuss the interactions that Replacement (Section 1.1) and Unrestricted Comprehension (Section 1.2) have with mathematical practice. Concerning Replacement we make it furthermore plausible that its status

[7] The idea of "finer resolution" is discussed in Subsection 3.2.3.

[8] Incurvati [54] is a book-length discussions of conceptions of set in which discussions about Replacement and the paradoxes feature prominently. I isolate several explanatory qualities in Section 3.1.

carries implications for set theory's ability to properly fulfill the foundational roles that Maddy [77] has identified.

To make progress in assessing the value of Linnebo's potentialist conception of set, we need to develop a general theory of conceptualization in mathematics. In Chapter 2, I will introduce a Lakatosian account of conceptualization, based on work by Incurvati [54], and explain on this account how conceptions differ from concepts. I will argue in particular that conceptions arise when an aspect which is subject to open-texture is sharpened in a concept that has a *foundational* subject matter. We will see, however, that not every sharpening to a conception can be legitimate, and that careful control must be exercised over the kind of stipulations we allow in conception formation. This problem especially arises when we try to deal with viciously circular justification. The novel solution I propose is to derive conception-admissibility from *explanatory coherence*, which can be attained by answering why-questions.

The quantity and quality of such explanations also act in support of the correctness of the conception. In Chapter 3 we compare the iterative and limitation of size conceptions in this sense as a case study, and show that the limitation of size answer to why V is not a set leads to viscous explanatory regress. We relate this regress in turn to the failure of the limitation of size conception to meet a naturality condition formulated by Boolos [13]. Then we introduce Linnebo's potentialist account of the iterative conception and show that this *does* meet the naturality condition, and moreover possesses several other virtues which we propose that a good account of the set-theoretic paradoxes should have.

Finally, in Chapter 4, we revisit the universal set paradox to see whether Linnebo's potentialism has any explanatory value over the conception-based minimalist explanation recently put forward by Soysal [100]. She holds that V is not a set simply because its existence conflicts with the axioms supported by the iterative conception (such as Pair and Foundation), and this conception is correct. This explanation threatens to leave one of the primary explanatory advantages of Linnebo's considerable modal apparatus up in the air, because Soysal's reasoning applies even on the minimal view which holds that "the content of the iterative conception is exhausted by saying that it is the conception of set according to which sets are the objects that occur at one level or another of the cumulative hierarchy."[9] I will argue, however, that Soysal's minimalist answer to the why-question (why is V not a set?) is explanatorily unstable, in contrast to the solution provided Linnebo's potentialism. I will therefore conclude that the modal apparatus does genuine explanatory work, and so Linnebo's account must be preferred

[9]Incurvati [53], p. 82.

as more correct.

For the most part, the following chapters can be read independently.

Chapter 1
Set Theory in Mathematical Practice

We discuss the interactions that Replacement and Unrestricted Comprehension have with mathematical practice. Beginning with Replacement (Section 1.1), we first discuss the distinction between extrinsic and intrinsic considerations, and how these modes of justification interact with each other on the foundational level (Subsection 1.1.1). It is questionable whether Replacement is proof-theoretically necessary for ordinary mathematics (1.1.2), but I will argue that its heuristic applications (1.1.3) convey a foundational significance which manifests as a method among the "MOTLEY of techniques of proof" that shape mathematics (1.1.4).

In Section 1.2 we turn to Naive Comprehension. This principle, which was lifted post hoc from Cantorian set theory, runs into several paradoxes. Cantor was aware of these problems, yet he did not perceive them as fatal to his ideas. He maintained that the implicated infinities are in a sense absolutely unlimited, and therefore cannot be constrained by mathematical bounds (1.2.1). However, Cantor's conception of this absolute infinite was nebulous, prompting some of his critics to find overarching mathematico-philosophical faults in the practice of mathematics to explain the inconsistencies (1.2.2). I will defend the position that the remedies brought forth by these skeptical programs, such as finitism and predicativism, detract from the richness of mathematical practice and, as such, should be rejected (1.2.3).

1.1 Replacement

1.1.1 Extrinsic and Intrinsic Justification

About the derivability of Replacement on the iterative conception, Boolos remarks:

> We do not believe that the axioms of replacement [...] can be inferred from the iterative conception. [...] For that there are exactly ω_1 stages does not seem to be excluded by anything said in the rough description [of the iterative conception]. (Boolos [12], pp. 227-8)

The problem that Boolos touches on here is that a conceptual derivation of Replacement requires an analysis of how the iterative conception internalizes that the

height of the set-theoretic hierarchy which it describes is a regular limit cardinal (and not say ω_1), without smuggling in "further thought[s]" that would beg the question about Replacement.[1] Typically these further thoughts attempt to work out the idea that the cumulative hierarchy extends 'as far as possible', which if not intrinsic to the rudimentary idea of the iterative conception, certainly harmonizes well with it. The added value of such accounts is disputed, but arguments along these lines are still researched.[2] Linnebo's modal recasting of the reflection principle ($\Diamond$-Refl) can also be seen in this context. We will take this up in Section 3.2.

Setting aside this *intrinsic* mode of justification, Boolos suggests to focus instead on *extrinsic* evidence to support Replacement:

> [Yet], although [the axioms of replacement] are not derived from [the iterative conception], the reason for adopting [them] is quite simple: they have many desirable consequences and (apparently) no undesirable ones. (Boolos [12], p. 229)

If we can find such extrinsic evidence for Replacement, it would demonstrate that the answers to questions 1 and 2 of Problem 1 impact mathematical practice.[3]

The converse implication from mathematical consequences to foundations is also possible, as we generally aim to offer intrinsic justifications for whatever we are extrinsically committed to. This foundational desideratum places demands on the conceptual underpinnings of our set theory, namely that it should be possible to derive the desired complementary intrinsic justifications from them. Linnebo [69] [68] can also be understood to argue in this spirit. Leveraging the need to explain Replacement and the paradoxes, he advocates for potentialism about the set-theoretic hierarchy. He argues that such a perspective provides the granularity to express novel "natural principles" within the framework of modal plural set theory, and contends that these can deliver the needed explanations.

The foundational quality conferred by fulfilling this desideratum lies in that intrinsic derivations are commonly deemed more robust than their extrinsic counterparts. The former reflects directly on the nature of set, and so is more intimately

[1] This is especially difficult because the required number of iterations far exceeds the first $\beth$-fixedpoint, and so is certainly beyond what anyone can capture in pictorial imagination (recall that $\beth_{\alpha+1} = 2^{\beth_\alpha}$ and $\beth_\lambda = \bigcup_{\alpha<\lambda} \beth_\alpha$). Boolos [14] (a later paper) is already skeptical about the existence of such smaller "pretty big" cardinals.

[2] See e.g. Incurvati [54], pp. 107-117.

[3] Barton et al. [3] have recently argued that the intrinsic/extrinsic distinction in justification is intractable and cannot be properly demarcated. I weigh their considerations in Subsection 2.4.5 in the context of the theory of conceptualization I develop in Chapter 2.

connected to the concept than any heuristic or pragmatic choice can be.[4] After all, any piece of extrinsic evidence, despite all the sound and fury it produces among practitioners, may ultimately signify nothing conceptual. Moreover, even if such evidence is not a fluke, it still does not automatically follow that it is indicative of the principle it is presupposed to be evidence for. Potter ([91], p. 388-9) furthermore argues that regressive (extrinsic) methods do not evidence the consistency of the principles that they are supposed to justify, nor can account for mathematical truth in the face of competing consequences in ordinary mathematics of alternative regressively argued for axioms.

1.1.2 Is Replacement Necessary?

The extent to which the discussion about Replacement affects mathematical practice is a much debated issue.[5] Scott-Potter set theory **SP** is a formalization of the iterative conception that lacks full Replacement, yet retains sufficient power to formalize (most of) classical mathematics.[6] On the other hand, Friedman (1971) has famously shown that Martin's Borel Determinacy Theorem (1975) is not formally provable in **SP**, despite that the theory of infinite games "no doubt deserves to be regarded as a genuine branch of mathematics, rather than merely within set theory itself."[7] Moreover, there are also instances of theorems in "ordinary mathematics" that do not involve determinacy, but which are still not provable in **SP**, such as Friedman's example "Every symmetric Borel set in the plane contains or is disjoint from the graph of a Borel function."[8] Nevertheless, such theorems are often considered *rarae aves*:

> I stand by my claim: I do not believe that explicit uses of replacement are going to seep through into more and more areas of pure mathematics in the way that uses of the axiom of choice became pervasive in the first half of this century. (Potter [90], p. 185)

It should furthermore be noted here that while Borel determinacy *provably* necessitates *some* Replacement, it does not require a lot. It is enough to assume

[4] See, e.g., Maddy [74], pp. 482-3.

[5] Kanamori [59], pp. 46-8.

[6] **SP** is equivalent to a slight strengthening of Zermelo set theory **Z** (Incurvati [54], pp. 48-49). The canonical model of these theories is $V_{\omega+\omega}$, but **SP** internally uses Scott ordinals, which do not rely on Replacement in the way that von Neumann ordinals do. For example, modifying the proof of Hartogs's Theorem, you can prove the existence of ω_1 in **SP** (but of course not of V_{ω_1}).

[7] Potter [89], p. 221.

[8] Friedman [38] discusses this and several other examples.

that $\beth_\alpha$ exists for all α, which over Zermelo's weak set theory **Z** does not even imply the existence of a $\beth$-fixedpoint.[9]

1.1.3 The Heuristic Use of Replacement

Notwithstanding these concerns about how much Replacement is truly necessary in the strict proof-theoretic sense, there are broader extrinsic reasons, coming from mathematical practice, to accept it. The use of Replacement is so ubiquitous and casual that it has become a staple in the mathematician's conceptual toolkit. For instance, when she mentions the set $\{\{n\} : n \in \mathbb{N}\}$, she refers to it *as such*. The commonplaceness of resemblant notation suggests that mathematicians really do think of replacing in $\mathbb{N}$ each n for $\{n\}$. They evidently consider this to be a useful and legitimate mathematical operation, even though in this and most other cases there are less natural alternatives, such as

$$\{x \in \mathcal{P}(\mathbb{N}) : \exists n \in \mathbb{N}\,(x = \{n\})\}\,,$$

which describes the same set without invoking Replacement.[10]

It may be objected that a convenient shorthand like this hardly presents an extrinsic argument for Replacement, and that, as in the above haphazard example, only some trivial attention to detail is required to avoid its use. However, this ignores that even when an instance of Replacement turns out to be technically eliminable, it can still function as a heuristic guide for the mathematical creative process. For example, Replacement can be used to unify many types of definitions and arguments by means of transfinite recursion (which is equivalent to Replacement over **Z**).

A concrete example of the guidance that Replacement can yield is found in Mary Rudin's 1971 solution to a 20 year old conjecture in topology about the non-existence of Dowker spaces.[11] To refute the conjecture she went beyond **SP** to construct a huge Dowker space of cardinality $\aleph_\omega^{\aleph_0}$.[12] This initial example put Dowker space theory on the agenda and spurred the construction of many more examples. Despite all this activity, however, it took until 1996 before Balogh [2] was able to show that smaller (continuum-sized) Dowker spaces can be constructed without Replacement.

---

[9] See for instance Leinster [66].

[10] For a slightly more extended argument along these lines, and an indication of how such notational questions connect to predicativism, see Avron [1].

[11] Kanamori [59], p. 81.

[12] Recall that $V_{\omega+\omega}$ is a model of **SP** but lacks such large sets. It would be of some interest to make explicit the precise assumptions on transfinite recursion (and thus Replacement) that Rudin's construction demands in her original paper, but this is beyond the scope of this thesis.

Rudin's earlier example thus provided a heuristic that influenced and guided further research, both mathematically and historically. It would therefore undoubtedly be mistaken to state in general that a posterior finding of a Replacement-free alternative to a proof or construction which uses it, *ipso facto* nullifies the force of the extrinsic argument derived from the original discovery. The technical insight leveraged by Rudin's actual construction, as well as the psychological and sociological precipitate of the knowledge that there is at least *some* Dowker space, suggested and directed research problems and strategies that would have otherwise likely remained unexplored. And at least some outgrowths of this research program are legitimate and interesting even to those who reject Replacement.[13]

A similar-spirited, but better known and non-set-theoretic example of this type of heuristics, is the guidance that (complex) analysis has provided to number theory. For instance, the theory of the analytically extended Riemann zeta-function provides a general and insightful framework that is able to track many problems in number theory.[14] In this case we can readily see that regardless of whether these problems individually have elementary solutions—such as Selberg and Erdős (1948) famously provided for the Prime Number Theorem, which was first proved in 1896 using zeta function theory—this does not detract from the significance of analytic tools in guiding the number theorist's intuition and practice. And this remains true even if we consider that the relevant complex analysis can technically, through certain codings, be embedded in number theory.[15] The heuristic significance of the analytic machinery provides in itself extrinsic justification for its use

[13]See Balogh's [2] introduction.

[14]See e.g. Kanamori [59], pp. 81-83, for a discussion. The "analytical extension" of a complex function is the unique differentiable function which extends it (where possible) to the complex plane.

[15]This is demonstrated in Takeuti [105], part II.

in number theory.[16]

1.1.4 The Foundational Role of Replacement

It is important, however, to draw a distinction between the eventual removal of Replacement (or complex analysis) as described above and other rather less profound kinds of eliminability. Friedman [37] gives an example of the latter "fake use" of axioms in the context of algebra: the often invoked theorem that every field has an algebraic closure requires some "serious set theory" to prove, in particular Zorn's Lemma. Even so, in any practical scenario, the general theorem can be so tailored that it holds for all relevant cases, without requiring such significant set-theoretic assumptions.

In cryptography, for example, someone might cite this theorem to obtain the algebraic closure of some concrete finite field. However, all considerable set-theoretic assumptions, like Zorn's Lemma, can be completely avoided in this case by straightforward adaptation of the proof of the general theorem to cover just the finite field instances. Moreover, the only reason why cryptographers and mathematicians do not usually bother with this, is because of the citational convenience of a single general theorem. This contrasts starkly with the above examples of Balogh's construction of a small Dowker space and the elementary Selberg-Erdős proof of the prime number theorem. There, the elimination of axioms, or conceptual tools more broadly construed, required the adoption of completely different concepts and techniques of proof. This, in contrast, manifests novel insight.

With respect to this foundational role that Replacement provides, Kanamori writes in its defense:

> The conceptual space provided by the set-theoretic universe as buttressed by Replacement has been increasingly brought into play in

[16] This relates to the function which Gödel assigned to extrinsic justification. In particular, he wrote that:

> even in case [the new axiom] 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.* The axioms for the system of real numbers, rejected by the intuitionists, have in this sense been verified to some extent owing to the fact that analytical number theory frequently allows us to prove number theoretical theorems which can subsequently be verified by elementary methods. (Gödel [39], p. 521, emphasis mine.)

> modern mathematics to generate examples and counterexamples from analysis to topology. (Kanamori [59], p. 81)

We might therefore plausibly claim, in the terminology of Maddy [77], that set theory requires some Replacement to fulfill its intended foundational duty of "Generous Arena" where "all of modern mathematics takes place side-by-side."[17] Without Replacement some genuinely mathematical and diverse proofs cannot be properly carried out, and their associated insights cannot be legitimately leveraged, despite the fact that the techniques and conceptualizations involved in a proof—what makes it *that* proof—seem at least as essential to mathematics as any particular constellation of raw theorems. This quality of contributing to the "MOTLEY of techniques of proof" that shape our mathematics, should thus be judged separately from any proof-theoretic necessity.[18]

Another extrinsic argument for Replacement that stems from foundational concerns is that it provides for uniform, implementation-insensitive mathematical construction. As noted by Kanamori:

> Proceeding in the cumulative hierarchy, the ranks $V_{\omega+1}$ and $V_{\omega+2}$ correspond via cardinality to reals and real functions, and this has advanced the presumption contra Replacement that classical mathematics, at least, can be accommodated at these low levels of the cumulative hierarchy. However, the roughest indifference to identification, that according to cardinal equivalence, has to be acknowledged and then affirmed by considerable coding and dexterity. (Kanamori [59], p. 79)

Forster [33] gives a powerful example of the phenomenon hinted at here. He emphasizes that set theory does not natively support many common mathematical constructions, but does have the ability to simulate them. The price paid for this flexibility, however, is that it introduces implementation-specific arbitrariness.[19] There are, for instance, many ways to implement an ordered pair. Depending on the circumstances, one or the other implementation might be preferable with respect to the technical or pedagogic details of some proof.[20] However, what you absolutely do not want is that your particular choice of coding influences

[17] Maddy [77], p. 298

[18] This quote is from Wittgenstein [113], p. 176. For further discussion of this topic, starting from this remark in the *Remarks on the Foundations of Mathematics* [113], see Hacking [49] pp. 48-51, 57-59.

[19] Forster [33], pp. 47-51.

[20] As an example, Hamkins [51] defines a pairing function which is "flat" in the sense that it is non rank-increasing. This choice of implementation has advantages in some set-theoretic arguments over the usual Kuratowski definition.

your theorems, but without Replacement this might happen: the existence of the Cartesian product $X \times Y$ for *arbitrary* definitions of ordered pairs implies Replacement.[21]

1.1.5 Conclusion

Through various illustrations, we have showcased the relationship between Replacement and mathematical practice. In view of this, it is imperative to provide an intrinsic justification for Replacement, based on our successful iterative conception of set. However, it is doubtful that this can be done on a minimalist account of the iterative conception, according to which "the content of the iterative conception is exhausted by saying that it is the conception of set according to which sets are the objects that occur at one level or another of the cumulative hierarchy."[22] In Section 3.2 we consider Linnebo's account of the iterative conception and briefly discuss how this might provide this intrinsic justification.

We now turn to the second example of interaction between foundations and mathematical practice related to the questions posed in Problem 1, namely the debate engendered by the paradoxes that arose in the naive set theory.

1.2 Naive Comprehension

1.2.1 Cantorian Set Theory and Paradox

The rapid rise of set theory as a powerful general framework for mathematics in the late nineteenth and early twentieth century prompted extensive scrutiny of its foundations. This effort led in particular to criticism of the indispensable role that actual infinities play in set theory, which went against the ancient contention that infinitudes should only ever appear as potential magnitudes. This potentialist conceptualization of infinity, rooted in Aristotle's *Physics*, had remained dominant until the nineteenth century. It was only during this period that this perspective slowly began to crumble under the pressure of the rapid abstractization of mathematics, propelled by the work of Dirichlet, Riemann, Dedekind and others.[23] Skepticism remained abound however. As late as 1906, Poincaré wrote

[21] Proof: let $(\cdot,\cdot)$ denote the standard pairing function and consider an arbitrary definable function F and set X. Now $(x,y)_F = (F(x),(x,y))$ also defines a pairing function and $F(X) = \{y \in \bigcup(X \times_F \{\emptyset\}) : \exists x \in X\,(y,(x,\emptyset))\}$, which requires only Separation and Union. This is Forster's example.

[22] Incurvati [53], p. 82.

[23] See Ferreiros [31] and Maddy [77], pp. 294-8.

that: "There is no actual infinite; the Cantorians forgot that, and they fell into contradiction."[24]

The Cantorian position was muddled in particular by the lack of a clear, published description of what constitutes being a set. In his main mathematico-philosophical investigation into the theory of the infinite (his Grundlagen), Cantor only wrote that:

> By a 'manifold' or 'set' I generally understand every multiplicity which can be conceived as one, i.e., every totality of definite elements which by means of a law can be bound up into a whole. (Cantor 1883, in Jané [58], pp. 391-2)

We might imagine here a notion of set-hood akin to Fregean extensions of concepts, in the sense that there corresponds a set to every definite property P so that the set contains just those objects that exemplify P.[25] If we assume that this logical separation of the universe suffices for collection into a set, we can conveniently define sets like $\{n : n \text{ is prime}\}$ and $\{x : x \text{ is a real}\}$, owing just to the logical form of these expressions and the definiteness of the involved predicates. Unfortunately, despite the strong intuitive appeal and uniformity of this *Unrestricted Comprehension* (UC) principle, the naive theory of set based on it is plagued by paradox. For example, we run into paradox when we consider the order-type of the totality of ordinals, or the cardinality of the cardinals:

Paradox 1 (The Burali-Forti Paradox)**.** By UC there is a set of all ordinals $\mathrm{Ord} = \{\alpha : \alpha \text{ is an ordinal}\}$. Because the ordinals are themselves well-ordered, there must also be an ordinal Ω representing the order-type $\mathrm{ot}(\mathrm{Ord})$ of Ord. But this means that $\Omega \in \mathrm{Ord}$, which implies $\Omega \subsetneq \mathrm{Ord}$. So $\Omega = \mathrm{ot}(\Omega) < \mathrm{ot}(\mathrm{Ord}) = \Omega$, and we have a contradiction.

Paradox 2 (Cantor's Paradox)**.** Let $C = \{\kappa : \kappa \text{ is a cardinal}\}$ be the set of all cardinals, which exists by UC, and let μ be the cardinality of $\bigcup C$. By Cantor's Theorem $2^\mu > \mu$ is also a cardinal in C, but then $\mu = |\bigcup C| \geq |2^\mu| = 2^\mu > \mu$, which is impossible.[26]

Pathologies can also occur when the membership criterion is apparently very simple, as in the set of all sets:

[24] See Feferman [29], pp. 6-9, about Poincaré's position. This quote is also borrowed from that paper.

[25] We say that a property is *definite* if it does not exhibit any ambiguity in application.

[26] Cantor's Theorem says that the cardinality κ of a set is strictly less than the cardinality 2^κ of its powerset. In other words, there is no injection of $\mathcal{P}(A)$ into A.

Paradox 3 (The Universal Set Paradox)**.** UC guarantees the existence of the universal set $V = \{x : x = x\}$ and its powerset $\mathcal{P}(V) = \{x : x \subset V\}$. By Cantor's Theorem $|V| < |\mathcal{P}(V)|$, but by universality also $\mathcal{P}(V) \subset V$. Hence $|\mathcal{P}(V)| \leq |V| < |\mathcal{P}(V)|$, which cannot be.

There are even instances of UC that generate paradox all by themselves, such as with the Russell set:

Paradox 4 (Russell's Paradox)**.** By UC there is a diagonalizing set $R = \{x : x \notin x\}$. If $R \in R$, then by definition $R \notin R$. If on the other hand $R \notin R$, then R satisfies the membership condition of R, so $R \in R$. Thus in both cases we end up in contradiction.

It would, however, be mistaken to ascribe to Cantor this *logical* conception of set. On the contrary, there is significant textual evidence that the paradoxes—quite fatal to Frege's project—did not affect Cantor's mathematico-philosophical outlook much at all.[27] For instance, commenting on Dedekind's use of "the totality of all things that can be objects of my thought" (a kind of V) in establishing that there are infinitely many things, Cantor writes to Hilbert:

> To my joy, I see from your esteemed letter that you recognize how important it should be to [Dedekind] the open publication of the foundations of my investigations on sets (foundations that you can find, clearly though intentionally somewhat hidden, in my 1883 *Grundlagen*, specially in the end notes); My foundations lie opposite to the core of his investigations, which is the naive assumption that all well-defined totalities or systems are always *consistent systems* as well. So you are also convinced that this assumption of Dedekind's is wrong. This I naturally saw just after the issue in 1887 of the first edition of his [*Was sind und was sollen die Zahlen?*] (Cantor 1889, quoted in Jané [58], p. 397, emphasis as in the original.)

No paradoxes arose from Cantor's conception of set, because he had always been careful to make a distinction between the proper infinite, which can always be increased further, and the *absolutely unlimited*, which is impossible to raise in quantity. Cantor took the extended number-sequence (the ordinals) as "an appropriate symbol of [this] absolute," because any stretch of numbers $\alpha, \beta, \gamma, \ldots$ that is bounded in a mathematically meaningful sense can be appended with a new limit number $\lambda > \alpha, \beta, \gamma, \ldots$.[28] The absoluteness of the number-sequence thus

[27] This point is defended in Jané [58], Menzel [79], and Lavine [65] (especially chapter 4).

[28] The quote is from Cantor's Grundlagen. It is found translated in Jané [58] p. 382.

consists in that this 'process' has no *mathematical* limit, and that the sequence therefore cannot exist as a completed entity that can be *determined* (assigned a number).

The Burali-Forti paradox is now neutralized, because on this view of the ordinals the pathology is merely a reflection of the absolute unlimitedness of the number-sequence. There simply cannot be a completed totality of all numbers since that would contradict their mathematical unboundedness. Thus, Cantor differed from the ancients primarily in that he denied that every actual infinite magnitude is *ipso facto* also absolute. This furthermore explains the cryptic distinction between different kinds of multiplicity that he put forth in a letter to Dedekind in 1899:

> [...] a multiplicity can be such that the assumption that *all* of its elements "are together" leads to a contradiction, so that it is impossible to conceive of the multiplicity as a unity, as "one finished thing". Such multiplicities I call *absolutely infinite* or *inconsistent multiplicities*.
>
> As we can readily see, the "totality of everything thinkable", for example, is such a multiplicity; later still other examples will turn up.
>
> If on the other hand the totality of the elements of a multiplicity can be thought of without contradiction as "being together", so that they can be gathered together into "*one* thing", I call it a *consistent multiplicity* or a "set". (Cantor 1899, in van Heijenoort [46], p. 114, emphasis as in the original.)

This distinction between consistent and inconsistent multiplicities is, however, in itself not very helpful in explaining the paradoxes. It just kicks the can further down the road. The ordinals do not form a set because they are absolutely infinite; but without a prior interpretation of the absolute, we can only say that the ordinals are absolutely infinite because if, on the contrary, they could be collected into "one thing," that would mean that they can be mathematically bounded, and this leads to contradiction, as we saw. The explanation we obtain from the Cantorian distinction between types of multiplicities then becomes in this case that the ordinals cannot form a set because the assumption that they can leads to contradiction. This does not provide any new insight.

What distinction Cantor did have in mind here is still contested. Analyses of his work variously suggest a theological, potentialist, or size distinction between the transfinite and the absolute. Some authors have even suggested that Cantor's conception of set was essentially iterative.[29]

[29] Jané [58] argues that the nature of Cantor's distinction between the bounded (sets) and the

The skeptics were however not swayed by Cantorian set theory. Kronecker is said to have mocked: "I don't know what predominates in Cantor's theory—philosophy or theology, but I am sure that there is no mathematics there."[30] In the next subsection we look at several skeptical interpretations of the paradoxes. I will argue that they adversely affect mathematical practice and hence should be avoided. In Chapter 3 we consider how the limitation of size conception measures against the iterative conception with regard to the paradoxes.

1.2.2 Skeptical Remedies

The discovery of the paradoxes incentivized skeptics like Kronecker and Poincaré to find overarching mathematico-philosophical faults in the practice of mathematics that could also explain the pathological nature of sets like V and Ord. The weaknesses so identified were traced by different schools of skeptics to various root causes, leading each school to impose their own specific limits on which principles and constructions to consider coherent and safe. This resulted in several fruitful analyses of the nature of mathematics beyond pure set theory, generalizing in some cases even past the strictly mathematical. Hence, the skeptical diagnoses carry substantial (prescriptive) implications for the practice of mathematicians. This line of research continues to this day.

We list some more radical examples of possible mathematico-philosophical culprits, diagnosing the source of the paradoxes in...

- (*Finitism*) the use of actual or completed infinities *at all*, thereby defying the giants of classical mathematics, who with Gauss protested against

> [...] the use of infinite magnitude as something completed, which in mathematics is never permissible. Infinity is merely a way of speaking, the real meaning being a limit which certain ratios approach indefinitely near, while others are permitted to increase without restriction.[31]

absolute changed from theological to potentialist between the publications of the Grundlagen and the Beitrage. Hallett [40] is a book length discussion of Cantor's work which puts his approach under limitation of size. Frápolli [36] argues against those that have interpreted Cantor's comments as implying some sort of iterative conception.

[30]This quote is attributed to Kronecker in the folklore. I was, however, unable to trace its origins, but it seems to be a fabrication (I cannot find the original German quote, for instance.) Nevertheless, it succinctly expresses the skeptical attitude towards Cantor's set theory in Kronecker's day.

[31]This quote is from a 1831 letter of Gauss to Schumacher. Linnebo and Shapiro [72], pp. 160-162.

- (*Predicativism*) the viciously circular definitions of the paradoxical sets. For instance, $V = \{x : x \text{ is a set}\}$ has to consider itself as a candidate for inclusion.[32]

- (*Dialetheism*) the mistaken assumption that a proposition and its negation cannot be both true at the same time without leading to explosion, and consequently trivialization.[33]

Of course, one could question why mathematicians should be disturbed by the perturbations of paradox within set theory *at all*, let alone consider such drastic revisionism. Practitioners can reasonably deny set theory any foundational priority, or even outright reject the need for foundations *simpliciter*. After all, mathematics is not really about sets, but rather about the numbers and structures that sets can be taken to model.

This point is well taken, and we shall not attempt in this thesis to trace out the precise foundational character of set theory. However, there are at least two reasons why this skepticism does not altogether undermine the significance of the paradoxes to mathematical practice.

First, set theory *does* provide indispensable foundational services to mathematics.[34] Against the "epistemic setbacks" of mathematics in the nineteenth century and onwards, set theory supplies mathematics with a universal language and versatile judge of proof. And it is *this* that allows for the coherent interoperability of ideal and abstract objects as they arise in various axiomatized systems, originating from such distinct mathematical fields as analysis and geometry. Therefore, because these services are currently not implemented in full by an alternative foundation, yet are crucial to the practice of modern mathematics, any shadow cast upon the set-theoretic background (as the paradoxes do) potentially also falls over its mathematical dependencies.[35] This is especially so since we know of "set-

[32]Predicativism originates with Russell and Poincaré. See Feferman [29], pp. 1-10, for a historical introduction.

[33]The use of dialetheism to explain the set-theoretic paradoxes is due to Graham Priest. See Incurvati [54], chapter 4 for an assessment of the position.

[34]Forster [33], p. 16, makes the point that "perhaps one would be better off trying to argue that set theory has a unifying rôle rather than a foundational rôle", and that "set theory as a single currency for mathematics is an easier idea to defend than set theory as a foundation for mathematics." We shall not attempt to delineate or argue the proper meaning of *foundations*, but there seems to be no particularly strong reason to restrict the term as Forster does to metaphysical and epistemological foundations. Instead, we use the term here liberally like Maddy ([77] p. 298), who identifies "Metaphysical Insight" and "Epistemic Source" as just two more (but mistaken) "foundational virtues."

[35]Maddy [77], pp. 294-301, 310.

theoretic assertions with reverberations through the whole of mathematics," like Replacement (as we argued) and the Axiom of Choice.[36]

The second reason is that besides its foundational role, set theory is also a *mathematical* theory of the infinite. So the paradoxes can plausibly be seen as signifying a problem with set theory's unrestricted ambition in this regard (as, e.g., Cantor did by separating sets from the absolute infinite). Where we should then draw the line of legitimacy with respect to rigorous reasoning about the set-theoretic infinities, in the light of the discovered limitations, is a discussion that touches upon infinitude generally as it occurs throughout mathematics. In this context, the finitist perspective on the paradoxes can be viewed as an extreme stance regarding the acceptability of infinitudes.

1.2.3 Mathematical Richness

Finitism, predicativism, and dialethism have all been developed into deep explanations of the paradoxes by their proponents.[37] However, while the exact strength of the systems that can be leveraged by these programs is a matter of debate, it is clear that they cannot close the gap with mathematical practice. Confinement to the legitimatized extent of these systems inevitably imposes far-reaching restrictions on our practice and impinges on the subject's progress.[38]

For example, a widely accepted argument due to Tait [104] puts the extent of first-order finitist arithmetic at provability in the very weak system known as Primitive Recursive Arithmetic (**PRA**).[39] Likewise, the proof-theoretic strength of predicative systems is generally considered to be bounded by the relatively small Feferman-Schütte ordinal Γ_0.[40] In terms of the second-order systems of arithmetic identified by the Reverse Mathematics program, this implies correspondences between finitism and the system $\mathbf{WKL_0}$, which is conservative over **PRA**, and pred-

[36]Forster [33], p. 16.

[37]Section 3.1 elaborates a little further on the meaning of depth intended here.

[38]This, of course, presupposes that we assign a certain significance to the practices of mathematicians. If one subscribes to a more holistic view of the justification of mathematical principles, which takes scientific needs as central, one might not be impressed by the arguments below about proof-theoretic strength. See also Feferman [29], pp. 21-3, about the "significance for the indispensability arguments of the positive developments of parts of mathematics by predicative means."

[39]Feferman ([30] and [29], pp. 24-27) presents a different argument that finitist arithmetic coincides with **PRA** based on Feferman's idea of "unfolding" a schematic system S to find "which operations and predicates—and which principles concerning them—ought to be accepted if one has accepted S."

[40]Feferman [29], pp. 15-16.

icativism and the system $\mathbf{ACA_0}$, which is conservative over **PA**.[41] To stress the weakness of these systems: $\mathbf{WKL_0}$ fails to prove the Bolzano–Weierstrass Theorem, and $\mathbf{ACA_0}$ cannot establish the Perfect Set Theorem. And, of course, neither system can prove any of the well-known, natural examples of arithmetical theorems of finite combinatorics independent of **PA**, such as the finite Ramsey Theorem involved in the Paris-Harrington Theorem, and Friedman's finite Kruskal's Tree Theorem.[42]

We, however, surely do not wish to abandon with the finitists and predicativists "the paradise which Cantor has created for us,"[43] and so we must contemplate less restrictive answers to quell the worry why some intuitively determinate collections, such as V and Ord, apparently do not consistently aggregate to form a totality as a set. This means that when we evaluate a proposed account of the paradoxes, we should at least check whether it satisfies the following criterion:

➢ MATHEMATICAL RICHNESS: an explication of the paradoxes should not depend on a mathematico-philosophical framework that demands impracticable modifications to the mathematical practice as it exists. It should not impinge on the subject's progress.[44]

1.2.4 Conclusion

We have seen in this section that our interpretation of what goes wrong with Unrestricted Comprehension exerts a strong influence on what we consider acceptable mathematics. In the upcoming Chapter 2 I present a general account of the notions of concept and conception which explains how the concept of set can be developed into various conceptions of set. These conceptions come with different diagnoses of, and remedies for, the paradoxes. In Chapter 3 I set up several other evaluative criteria besides MATHEMATICAL RICHNESS. As a case study, we then compare the iterative and limitation of size conceptions. I will argue that the latter's account of the paradoxes is defective according to the proposed criteria. We then look at Linnebo's potentialist view of the iterative conception, according to which the

[41] Feferman [29], pp. 19-21.

[42] Graham Priest's dialetheism, arguably presents even more of a shake-up to mathematical practice. See e.g. Incurvati [54], chapter 4. However, it should be noted that Beall (for example in [6]) has attempted to close the gap with mathematical practice using so-called "recapture" results. I do not have the time and space here to evaluate these attempts.

[43] Hilbert [48], p. 191.

[44] See Section 3.1 for the other criteria I propose that any satisfactory account of the set-theoretic paradoxes should meet.

set-theoretic pathologies originate with the conflation of the actual and potential infinite:

- (*Set-Theoretic Potentialism*) The full set-theoretic universe is a potential species that behaves much like Aristotle's potential infinite, except that it circumscribes numerous smaller actual infinities. Paradox arises when the universe is erroneously perceived as complete, whereas it can be extended indefinitely.

I will argue that Linnebo's account meets all the proposed evaluative criteria.

In Chapter 4 we consider whether this potentialism has any explanatory advantages regarding the paradoxes that would justify its considerable assumptions, and especially its modal apparatus, compared to the more lean "minimalist" view of the iterative conception that Soysal [100] advocates. We will draw for this analysis in particular on the framework we have developed in Chapter 2.

Chapter 2
Conceptions and the Significance of Why-Questions

In the last chapter we argued for the mathematical relevance of both Replacement and our diagnosis of the paradoxes associated with Unrestricted Comprehension. We also saw that conceptions of set may have widely different evaluations of these two principles. To sort this out we need to engage the following questions:

Question 1. *What are conceptions and how do they differ from concepts?*

Question 2. *How can we judge between conceptions? In other words, how does the notion of correctness manifest in conceptualization?*

Question 3. *How do we assess which axioms a conception sanctions?*

Any attempt at a theory of justification on a conception must surely first explain what conceptions *are*. Therefore we shall begin with Question 1 in Section 2.1. I will present an account of the notion of conception, building on Incurvati [54], according to which conceptions arise because of 'blanks' in concepts whose subject matter is under (at least) some pressure to be conceptually consolidated, but where the community of practitioners cannot (yet) make a decision. I connect this account of conceptualization to Waismann's idea of *open-texture* and, following Shapiro [97], show that this notion can be extended to mathematics on the Lakatosian view of mathematics as a "quasi-empirical" science which is in constant revolution (Subsection 2.1.1). These considerations then yield my proposed definition of a conception:

Definition (Conception)**.** A *conception* sharpens a concept by fixing an aspect of open-texture in the concept without breaking the conceptual continuity with regard to the subject matter.

This formulation highlights a crucial component of my interpretation of open-texture: the presence of open-texture not only requires us to decide how to sharpen the underlying concept, but it also mandates the ability to justify this decision, thereby establishing a notion of *correctness*. Without this possibility, the continuity between concept and conception would be lost, and the various conceptual

alternatives would collapse into distinct concepts altogether. So the question of why a concept should be sharpened in a particular way ought to make sense. In this context I will distinguish between *Fregean* concepts, like the concept of group, and *Lakatosian* concepts, such as the concept of set. I then argue that this distinction defines a spectrum property, namely *Lakatosianess*, which is related to the power of the underlying sense of correctness and subject matter continuity. I will argue that only sufficiently Lakatosian concepts admit conceptions.

In Subsection 2.1.2 I trace the origin of subject matter continuity to the role that a concept fulfills. I propose that the Lakatosianess of a mathematical concept relates to the degree to which a concept is designed to UNIFY, in the sense of Maddy [75]. Whether a concept is Lakatosian therefore depends on the roles assigned to the concept within a community of practitioners, in particular whether it is to serve a more *foundational* or a more *instrumental* purpose. This view harmonizes well with Tanswell's [106] recent argument that "mathematicians are a kind of conceptual engineers."

We return to the matter of conceptual correctness (Question 2) in Section 2.2, where it is also connected to Question 3. While we have considerable freedom in shaping our conceptions, each conception must come with some notion of correctness. I will show that this introduces the following dilemma: either we must justify when a conception is *admissible* or we need to provide direct explanations of conceptual correctness. However, Incurvati [54] has demonstrated that direct correctness arguments, even for sets, are fraught with difficulties. As a result, the latter horn of the dilemma proves intractable, leading us to focus on the Admissibility Problem (see Problem 2) in Subsection 2.2.1. I will argue that this problem hinges on the permissibility of *stipulation* in conception formation, especially regarding circular justification (Problem 3). We address this issue in Subsection 2.2.2.

Waxman [112] has argued that mathematicians have an epistemic freedom which is only contingent on "meeting minimal standards of logical *coherence*." Any restriction on stipulation should therefore be implied by these basic criteria. In Section 2.3 we examine two alternatives for such criteria for coherence. One option is to assert in hilbertian style that consistency suffices for existence and coherence (Subsection 2.3.1). Another, proposed recently by Isaacson [55], views coherence in terms of use in practice (Subsection 2.3.2). I will argue, however, that both proposals inadequately address the role of stipulation in conceptualization.

Finally, in Section 2.4, I present my case that coherence may instead be derived from explanation. I contend that our demand for coherence in foundational concepts is essentially a demand for the stability of our mathematical concep-

tual framework under the pressure of inconsistency, improper conceptualization, mathematico-philosophical incongruence, and failures of conceptual integration. In Subsection 2.4.1 I argue that explanation contributes to this stability and, moreover, that explanatoriness should itself be regarded as a virtue against which one can compare the correctness of conceptions. Hence we should strive for *explanatory saturation*, meaning that anything that *can* be explained *should* be explained, and as powerful as possible.

In Subsection 2.4.2 I connect these ideas to the old debate in epistemology between *coherentism* and *foundationalism*. Coherentism interprets justification holistically as a relation that is always expressed in reference to a "web of beliefs." When the mutual support furnished by this web takes the shape of intertwined explanations, we obtain *explanatory coherence*. I show that the standard arguments against coherentism pose no particular problem when we argue in this manner for the coherence of a system of mathematical conceptions. Furthermore, in Subsection 2.4.3 I hypothesize that explanation can be effectuated simply by answering *why-questions*. This view is supported by recent empirical work of Pease et al. [87].

We then return to the problem of circular justification which I outlined in Section 2.2, and the example case of justifying Infinity versus justifying measurable cardinals. I apply the why-question coherence framework to argue that the stipulation of Infinity is admissible, while the stipulation of measurables is not (Subsection 2.4.4). Support in favor of measurables, such as their utility and structural connections with axioms of determinacy, is tied to theories and not conceptions, and therefore does not effectuate explanatory coherence. This evidential distinction between theories and conceptions parallels the difference between extrinsic and intrinsic justification. Barton et al. [3], however, have recently challenged the tractability and demarcation of these justificatory categories, and so in Subsection 2.4.5 I will attempt to refute any criticism that may arise from their considerations. Regarding tractability I contend that their approach of prediction and confirmation fares no better in avoiding debate. And from the standpoint of a conceptual engineer there is no inherent issue with implementing conceptual changes to incorporate extrinsic evidence, so the demarcation problem is also mitigated.

2.1 Concepts and Conceptions

To understand what a conception is and what role it plays we must first distinguish it from the notion of a concept.[1] We do not need to take a particular stance on

[1] Our use of the terms "concept" and "conception" is not necessarily standard in the literature, but inspired by Incurvati [54], chapter 1.

what concepts are, important here is only that concepts (as we use them here) are *discrete*: concepts C_1 and C_2 are either identical or they describe distinct ideas that can in no way be seen as rival accounts of the same topic (strictly bounded away from each other). So whenever two speakers disagree on the characterization of a concept C, either at least one of them must be deemed to not fully possess C, or the characterizations in question must refer to conceptually distinct notions that have been erroneously grouped under the same concept term. There is thus no pressure whatsoever to isolate a correct notion among some concepts, and in fact there is not even a sense of resolution that possibly might be applied between them.

As an example, the concept of set is the rudimentary idea of a bunch of well-differentiated things collected together into a new object—something like a 'bag' or 'container'[2]—which is completely determined by its members. Formulating an explicit set concept like this helps distinguish between several distinct but related notions that stem from the same inchoate jumble of thoughts connected to the general idea of collection. Drop unique decomposition and you get fusions, drop singularity and you get pluralities, drop differentiability and you get multisets, etc. These are however simply different concepts of collection and it is nonsensical to try to compare them against each other or against set.[3]

A *conception* of a concept C extends C with an interpretation of some of the 'blanks' left underdetermined by C *without this amounting to changing the topic*. So when speakers discourse about some concept C, but disagree on some conceptual aspect about which there is no consensus among the wider community of users of C, we say that the speakers have different conceptions of C. Between conceptions the subject matter is thus more *continuous* than between concepts—namely the conceptions are about whatever C is about—and therefore there is at least some pressure for them to be consolidated. For example, speakers who are considered by some language community to be competent users of the term "fair" may still disagree about the fairness of a particular situation. Warring tribes will certainly disagree about the fairness of the other party, yet it would be strange to suggest that this must mean that one of the tribes simply misunderstands what it means to be fair (certainly they will think *they* are fair!). There is also pressure to find common ground and ameliorate differences to arrive at a solution that both sides can agree is fair.

Incurvati [54] takes a conception based on a concept C to be an answer to

[2]See, e.g., section 3 of Oliver & Smiley [83].

[3]To abbreviate "the concept of X" we will sometimes write X to avoid confusion with the things called X. Thus set refers to the concept of set.

the question of "What [it is] to be something falling under C" that goes beyond, but not against, C. This definition seems to only acknowledge *delineative* differences (roughly, extensional distinction), which makes sense on Incurvati's account, because he defends a brand of minimalism that purports to rid us of any metaphysical conundrums in our conception of set (see Section 4.4). On the other hand, Incurvati also argues that some of the appeal of the iterative conception is due to "the fact that it provides a convincing explanation of the paradoxes."[4] To assess Linnebo's potentialist project [69] [68] we need to compare conceptions of set that rely on different logics to treat the sets with, but which are yet co-extensional. This requires that we recognize that explanation may be an aspect of conceptualization, otherwise the whole project would become moot. Therefore, we assume that the blanks filled in by a conception can also be *explanative* in nature (roughly, intensional distinction).

In this way we *develop* a concept into a conception—or alternatively we say that a conception is *formed* from a concept—as a more refined account of that concept. There is however no sharp boundary between what is part of a concept C and what is part of some conception of C.[5] For instance, in the early days of set theory there was some debate over whether sets should be considered an intensional or an extensional species. The idea that the way in which a set arises is inconsequential for its criterion of identity has become part of the concept of set only later. If someone now disputes the extensionality of sets and proposes an intensional conceptualization, we will dismiss it as a change of topic.

2.1.1 Open-Texture

One may object against this concept versus conception distinction that it can only arise in such ambiguous contexts as human social life, but that in science, and especially in mathematics, our concepts are precisely delimited and do not contain any 'blanks' to be filled in. Thus there are no conceptions in mathematics, just inadequate concepts that need to be replaced, or so the objection goes. However, it was recognized by Waismann that this is typically not the case and that scientific concepts admit *open-texture*:

> The notion of gold seems to be defined with absolute precision, say by the spectrum of gold with its characteristic lines. Now what would you say if a substance was discovered that looked like gold, satisfied all

[4] Incurvati [53], p. 84.

[5] This preempts some Quinean worries that might arise from the similarity of the concept versus conception distinction to the analytic/synthetic distinction.

> the chemical tests for gold, whilst it emitted a new sort of radiation? [...] Try as we may, no concept is limited in such a way that there is no room for any doubt. We introduce a concept and limit it in *some* directions; for instance we define gold in contrast to some other metals such as alloys. This suffices for our present needs, and we do not probe any farther. We tend to *overlook* the fact that there are always other directions in which the concept has not been defined [...] we could easily imagine conditions which would necessitate new limitations. In short, it is not possible to define a concept like gold with absolute precision; i.e., in such a way that every nook and cranny is blocked against entry of doubt. That is what is meant by the open texture of a concept. (Waismann [110], p. 120, emphasis as in the original.)

The point is that there is no fact of the matter whether or not the oddly radiating lump of gold-like material is *in fact* gold. The case is not fixed by our current concept of gold, as it falls outside the standard domain of application. Another example that Waismann entertains is a cat which grows to a gigantic size or shows other queer behavior not normally found with cats. He then asks: "Shall I, in such a case, say that a new species has come into being? Or that it was a cat with extraordinary properties?"[6] Once again, Waismann argues that a concept such as cat that meets our current language needs is not as precisely delimited as we might think, and that non-standard examples are in fact not hard to come by. In general, we will say that a concept C admits open-texture whenever there are cases for which a competent user of C can reasonably assert either that C applies or that it disapplies.[7]

These examples show that empirical terms from science and everyday language are subject to open-texture, and that as a result their meaning is continuously evolving in tandem with ongoing empirical observations. Waismann describes this account of language in flux as "the essential incompleteness of an empirical description."[8] In mathematics, however, we seem to have a much greater degree of control over our concepts, as is evidenced by the rigor with which practitioners usually operate. We can therefore maintain that mathematical and empirical con-

[6] Waismann [110], p. 119.

[7] This definition of open-texture due to Shapiro is somewhat broader than Waismann's because it also considers borderline cases (in the sense of vagueness) that do fall under the standard domain of application. For example, normal human heads are part of the domain for the concept of baldness, yet it may not be fixed by baldness whether some borderline bald head is bald or not. Conversely, when there is no fact of the matter whether a concept applies or disapplies, it seems that we should be able to settle the matter either way. See Tanswell [106], pp. 4-8.

[8] Waismann [110], p. 121.

cepts are fundamentally different, so that the above objection against the concept versus conception distinction can be upheld. According to Waismann as well, we reach absolute precision in mathematics, and its concepts therefore do not admit open-texture:

> Goldbach's hypothesis [...] may be undecidable [...] But this in no way detracts from the *closed* texture of the mathematical concepts. If there is no such thing as the (always present) possibility of the emergence of something new, there could be nothing like the open texture of concepts. (Waismann [110], pp. 123-124, emphasis as in the original.)

The problem with this dogmatist position, however, is that it is refuted by the actual history of mathematics. In his well-known case study *Proofs and Refutations* [64], Lakatos presents a rational reconstruction of the history of Euler's formula $V - E + F = 2$ for polyhedra, paying special attention to the concepts involved, such as polyhedron, face, and vertex.[9] He demonstrates that these concepts have historically not been exclusively shaped by their pre-theoretic meaning in ordinary language. Rather, the proofs *about* these concepts have also been pivotal in their evolution, along with the critiques and counterexamples that emerged alongside these proofs. The driving force behind conceptualization in mathematics, Lakatos concludes, is therefore primarily our desire to have concepts that are fruitful and about which we can prove interesting results. Consequently, mathematical concepts are not merely static entities that are passed down to us once and for all with absolute precision, as the Fregean tradition understands it.[10] There is ample room for dynamic interplay between our concepts and what we discover about them.

Lakatos skillfully illustrates this dialectical component of mathematical conceptualization through a classroom dialogue. The teacher opens with a presentation of Cauchy's proof of Euler's formula, which involves imagining a given polyhedron as a hollow skeleton of vertices and edges covered with a thin rubber surface, and removing one of its faces to stretch the remaining surface flat. This proof idea prompts several rounds of criticism and counterexamples from the students, and a matching series of reformulations or refinements of Euler's conjecture and the involved concepts. The students propose, for example, a cube with a cube-shaped hole in one of its faces, a star-polyhedron consisting of 12 star-pentagons, a "picture frame" cube that has a tunnel through its center, and a cube with a cube-shaped cavity. Some of these refutations are *global* in nature and violate the

[9] Here V represents the number of vertices, E the number of edges, and F the number of faces in a given polyhedron.

[10] For a comparison between the Lakatosian and Fregean approaches see Schlimm [95].

conjecture $V - E + F = 2$ itself, others agree with the formula but 'fail' *locally* in the sense that they violate some (hidden) lemma used in Cauchy's proof.

Various types of maneuvers to alleviate such local and global tensions are discussed, and in the spirit of his rational reconstruction, Lakatos traces these patterns of dynamic interplay largely along historical lines. One such maneuver is to insist that the proposed global counterexamples are merely pathological cases which obviously do not fall under polyhedron.[11] Alternatively, these "monsters" can also be interpreted as arising from a surreptitious *expansion* of the naive concept of polyhedron, which is inadvertently stretched in our attempts to extract the maximum possible content from Cauchy's proof.[12] Another line to save the conjecture is to "withdraw strategically" and "draw the borderline modestly, but safely: *All convex polyhedra are Eulerian.*"[13] However this ignores a whole class of local counterexamples as there are many concave polyhedra that do fit Euler's formula. We can also go the opposite route and reject Cauchy's proof. We may then either hold that Euler's conjecture is only a step in the right direction of the proper generalization, or conclude that the notion of polyhedron is too theoretically fragmented for there to be any interesting results about.

The point of Lakatos's historical analysis is to demonstrate that mathematics is in a permanent state of revolution, and that in this sense it is much closer to scientific practice than is commonly assumed, including by Waismann. The case study of polyhedron shows that, at least in some cases, mathematical theories are subject to the above-mentioned "quasi-empirical" patterns.[14] Navigating all the possible mixtures in which these patterns can be coalesced necessarily involves a

[11] Lakatos dubs this the method of "monster-barring":

> STUDENT DELTA: But why accept the counterexample? We proved our conjecture—now it is a theorem. I admit that it clashes with this so-called 'counterexample'. One of them has to give way. But why should the theorem give way, when it has been proved? It is the 'criticism' that should retreat. It is fake criticism. This pair of nested cubes is not a polyhedron at all. It is a monster, a pathological case, not a counterexample. (Lakatos [64], p. 15)

[12] This "refutation by concept stretching" provides an alternative interpretation of monster-barring:

> STUDENT PI: [...] In [the refutationist's] critical zeal they stretched the concept of polyhedron, to cover objects that were alien to the intended interpretation. The conjecture was true in its intended interpretation, it was only false in an unintended interpretation smuggled in by the refutationists. (Lakatos [64], p. 90)

[13] Lakatos [64], p. 30, emphasis as in the original.

[14] See in particular Lakatos [63], section 2.

great deal of choice, and so at some point in time there was simply no fact of the matter about whether or not a picture frame should be classified as a polyhedron. In other words,

> The notion of polyhedron exhibited what Waismann calls open-texture. This open-texture did not prevent mathematicians from working with the notion, and proving things about polyhedra. Still, at the time, it simply was not determinate whether a picture frame counts as a polyhedron. [...] When the case did come up, and threatened to undermine a lovely generalization discovered by the great Euler, a decision had to be made. (Shapiro [97], p. 438)

Moreover, open-textured concepts are not only found in mathematical disciplines which depend on spatial intuition. Shapiro [97] shows, for instance, that similar interactions between concepts, conjectures, and proofs, took place in the emerging field of recursion theory between 1930 and 1960, giving rise to the modern Church-Turing notion of computability:

> The concept [of computability] was not delineated with enough precision to decide every possible consideration concerning [the allowable] tools and limitations. And just as with Lakatos's example of a polyhedron, the mathematical work, notably Turing's 1936 argument and the efforts of the other founders [...] sharpened the notion to what we have today. In other words, the mathematical work served to set the parameters and thus sharpen the original pre-theoretic notion. (Shapiro [97], p. 441)

Now that we have introduced and defended the notion of open-texture in mathematical concepts, we can reaffirm our concept versus conception distinction in terms of this idea:

Definition 1 (Conception)**.** A *conception* sharpens a concept by fixing an aspect of open-texture in the concept without breaking the conceptual continuity with regard to the subject matter.

This formulation highlights an aspect of open-texture that is not apparent in Waismann's empirical examples, but which I think is crucial, namely that a sharpening requires not only that some choice be made, but also that this choice can in principle be justified. That is, that there is ground for a notion of *correctness*. Without this possibility, the sense of continuity between concept and conception would

be lost, and the various conceptual alternatives would just collapse to different concepts altogether.

The concept of group, for instance, leaves it open whether a group is commutative or not, but commutativity is clearly not a matter of open-texture in group.[15] I hold that *open-texture demands that the question of why it should be sharpened in a particular way, and not another, makes sense*, and whether this is the case depends on the nature of the underlying concept. Field and ring are on my view not conceptions of group—although they likewise extend the concept—because the purely structural role that we have assigned to group does not come with any sense of correctness that could possibly decide, even only in principle, between these concepts. Adopting the terminology of Schlimm [95], we may say that group is strictly *Fregean*, whereas polyhedron would be at least partially *Lakatosian*. Schlimm makes this distinction thus:

> According to the first [notion of mathematical concept], concepts are definite and fixed; in contrast, according to the second notion they are open and subject to modifications. I will refer to these two notions as 'Fregean' and 'Lakatosian', because they are based on ideas that have been emphasized by Frege and Lakatos. (Schlimm [95], p. 127)

In conclusion, I understand *Lakatosianess* as a spectrum property which is related to the power of the underlying sense of correctness, with Fregean concepts on one extreme (no sense of conception formation or correctness at all), and concepts with perfect continuity of subject matter on the other (at most one correct sharpening possible).[16] For instance, as a matter of observation of our mathematical practice, there is high pressure to resolve conflicts between different conceptions of set, moderate pressure to revolve conflicts between different conceptions of computability, and no pressure whatsoever to resolve 'conflict' between group and field.

When a concept is sufficiently Lakatosian it may admit conceptions. This is consistent with our earlier observation that we cannot draw a hard line between the notions of concept and conception. There are conceptions of set, and there are no conceptions of group. But for computability we have an intermediate mixture of concept-splitting and conception-formation. For instance, the sense of computability in the Church-Turing thesis (CT) does not compete with the idea of

[15] Shapiro told me in conversation that on his view commutativity may be regarded as a matter of open-texture. So I clearly diverge from Shapiro here.

[16] Again, this is definitely not Shapiro's view.

computability in the Cobham–Edmonds thesis.[17] But it does compete, for example, with the notion that Kalmár put forward in his plausibility argument against CT, which makes it possible to 'compute' the improper minimalization of any computable two-place function $\varphi(p, q)$ by the following 'method':[18]

> Calculate in succession the values $\varphi(p, 0)$, $\varphi(p, 1)$, $\varphi(p, 2)$, ... and simultaneously try to prove by all correct means that none of them equals 0, until we find either a (least) natural number q for which $\varphi(p, q) = 0$ or a proof of the proposition stating that no natural number y with $\varphi(p, y) = 0$ exists; and consider in the first case this q, in the second case 0 as result of the calculation. (Kalmar[19])

2.1.2 Foundational and Instrumental Concepts

Let us summarize the previous section: Lakatosian concepts arise when there is a sense of continuity of subject matter between a concept and (some of) the ideas that build on it. When this continuity is strong enough, we refer to these different ideas as conceptions of the underlying concept, and call the blanks that they fill in open-texture. Due to the assumed continuity of subject matter, conceptions must allow for some notion of conceptual correctness that can be applied to select the right conception, at least in principle. The reason for this is that without such a criterion it would not be possible to distinguish between concepts and conceptions at all. Hence, if we are to make progress on the questions posed at the outset of this chapter, we must understand, before anything else, how subject matter continuity emerges. Why is group Fregean, but set Lakatosian?

The research of Maddy [75], [76], [77] on the foundations of mathematics suggests that this continuity may originate from the (mathematical) roles that a concept is intended to serve:

> Now let's return to the purported analogy between group theory and set theory: trying to settle CH is like trying to decide if a group must

[17] The Cobham–Edmonds thesis in computational complexity says that a language can be decided *efficiently* if it can be decided in time $\mathcal{O}(n^k)$ for some $k \in \mathbb{N}$ (i.e. polynomial time).

[18] The improper minimalization of the partial two-place function f is the one-place partial function g, where $g(p)$ is defined to be the least y such that $f(p, y) = 0$ if there is such a y, and 0 otherwise. Not all improper minimalizations of Turing-computable functions are Turing-computable. For instance, the improper minimalization of the Turing-computable function $f(x, y)$ whose value is 0 if y is the code of a complete computation of the Turing machine whose code is x, and whose value is 1 otherwise, solves the Halting problem.

[19] See Shapiro [97], pp. 447-9, for a discussion of Kalmár's argument and its relation to the presence of open-texture in the concept of computability. The above quote of Kalmár is taken from Shapiro's paper.

> be commutative; both theories have many different models, some of which satisfy CH or the commutative law and some of which don't; there's no reason to try to select out one of these models as 'right' or 'intended'. If we now reconsider this argument while bearing in mind the contrasting goals of set theory and group theory, it no longer sounds so persuasive. [...] Group theory is *designed* to bring together a wide range of disparate mathematical structures, it would make no sense to try to rule out some of those structures as groups. [...] On the other hand, set theory is [...] *designed* to provide a foundation for classical mathematics, to provide a single arena for mathematical existence and mathematical proof, [so] it does make sense to try to make our theory of sets as decisive as possible, to try to choose between alternative axioms, to try to rule out models that do this foundational job less well than others, in short, to UNIFY. There's nothing wrong with group theory or set theory, but they are aimed at different mathematical goals. (Maddy [75], pp. 21-22)

So we have *foundational* concepts like set on the one hand and, let's say, *instrumental* concepts such as group on the other.[20] I hypothesize that the Lakatosianess of a concept relates to the degree it is subject to this unifying foundationality, and thus depends at least in part on the roles assigned to the concept within a community of practitioners. I propose the following axes of foundationality along which Lakatosianess can manifest (but there may be more):

- **Ontological independence**: the degree to which a concept is responsible for its own ontology. For example, whether there is a group that is a counterexample to the Whitehead problem depends on our set-theoretic background assumptions, and *not* on our concept of group. But whether there is a universal set is completely determined by our theory of set.

- **Conceptual integration**: the level of interconnectedness of a concept with other concepts. For instance, our beliefs about the (ordinal) numbers may influence our beliefs about sets, and vice versa, but this is not at all the case for groups versus fields.

[20] This foundational versus instrumental concept distinction is not concerned with the mathematical *significance* of an idea, as in that "[the idea] can be connected, in a natural and illuminating way, with a large complex of other mathematical ideas." (Hardy [43], p. 90) Group is certainly significant in Hardy's sense, yet instrumental.

- **Pre-theoreticalness**: whether or not a concept is fixed in some directions prior to theorizing.[21]

For example: group checks none of these criteria, polyhedron at least checks pre-theoreticalness, computability plausibly checks this and conceptual integration, and surely set checks all three.

What is at stake here between instrumental and foundational concepts is reminiscent of the debate that took place near the end the nineteenth century between proponents of the new axiomatic/formal method and the old propositionalist vanguard. Frege embodied the old approach and held against Hilbert—an icon of the new method—that "it can never be the purpose of axioms and theorems to establish the reference of a sign or word occurring in them; rather, this reference must already be established." Hilbert replied:

> You write "...From the fact that axioms are true it follows that they do not contradict one another." I was extremely interested to read just this proposition in your letter, because for as long as I have been thinking, writing, and lecturing about such things, I have always been saying the opposite: If the arbitrarily posited axioms together with all their consequences do not contradict one another, then they are true *and the things defined by these axioms exists.* For me, this is the criterion of truth and existence.[22]

Hilbert's viewpoint here, which has been sloganized as "existence is consistency," came to the foreground during a period of rapid algebraization of mathematics in the decades around the turn of the twentieth century. During this period emphasis shifted from structures such as *the* natural numbers and *the* Euclidean

[21] Pre-theoreticalness provides grounds for Lakatos's notion of a *heuristic falsifier*:

> The potential falsifiers of science, roughly speaking, express the 'hard facts'. But is there anything analogous to 'hard facts' in mathematics? [...] if we insist that a formal theory should be the formalisation of some informal theory, then a formal theory may be said to be 'refuted' if one of its theorems is negated by the corresponding theorem of the informal theory. Not all formal mathematical theories are in equal danger of heuristic refutation in a given period. For instance, *elementary group theory* is scarcely in any danger [...] heuristic refutations seem to be inconceivable. (Lakatos [63], pp. 213-4, emphasis as in the original.)

Some examples of heuristics falsifiers were discussed in the previous subsection in the context of polyhedron.

[22] These quotes are referenced in Isaacson [55], pp. 7-8, where he discusses the Frege-Hilbert debate. The emphasis here is mine. The debate is also analyzed in Dean [26], Blanchette [10], and Coffa [24], chapter 7.

plane, to structures like groups, fields, and systems of geometry (in particular hyperbolic and spherical geometry). But as fruitful as such algebraization has been in the history of mathematics, the radical interpretation of equating existence to consistency is at odds with our experience of foundational concepts. The Second Incompleteness Theorem implies the consistency of $\mathbf{PA}+\neg\mathrm{Con}(\mathbf{PA})$, provided **PA** is consistent, but this is not a coherent theory of some pluralistic natural number concept. We understand the concept of number foundationally, in reference to the particular pre-theoretical structure $\mathbb{N}$ of which we accept **PA**. Therefore any sort of 'number' that codes a proof of $\neg\mathrm{Con}(\mathbf{PA})$ must conceptually (and meta-theoretically) not really be a number, despite that the "existence is consistency" criterion guarantees it to exist as an object of the *arithmetical* theory $\mathbf{PA}+\neg\mathrm{Con}(\mathbf{PA})$.[23]

Likewise, Frege's position clashes with the actual development of mathematics in the twentieth century, which did increasingly emphasize mathematical objects (especially in geometry) as things understood through their defining axioms, without conceptual content beyond what can be grasped up to isomorphism. Our Maddy-inspired analysis suggests that the Frege-Hilbert debate turns on the failure to appreciate the importance of the role of a mathematical concept in determining its nature. The variability of this role implies that asking what the nature of mathematical concepts is does not constitute a well-defined question, so the

[23]It must be noted that Hilbert's views matured significantly between the publication of *Grundlagen der Geometrie* (1899), which spawned the Frege-Hilbert debate, and the release of his and Bernays's first volume of *Grundlagen der Mathematik* (1934). In fact, in the latter work Hilbert and Bernays introduce a distinction that overlaps significantly with the instrumental/foundational concept distinction presented here, although it (of course) does not recognize the connection I claim to Lakatosianess, open-texture, and the active role that practitioners play in the nature of conceptualization. Dean [26] summarizes that

> by the time of Grundlagen der Mathematik, [Hilbert] and Bernays had also come to systematically distinguish between what they called "formal axiomatics" and "contentual axiomatics". The former applies to geometrical theories [...] whose consistency cannot be secured by exhibition and whose axioms thus not only lack "a special epistemic relation to [a] specific subject matter" but also "transcend the realm of experience and intuitive self-evidence" [...] By contrast, Hilbert and Bernays describe contentual axiomatics as being grounded "in common experience [which] presents its first principles either as self-evident facts or formulates them as extracts from experience-complexes." And it is precisely this sort of experience which they took to ground elementary number theory. (Dean [26], pp. 377-8.)

Doherty [27] presents a nuanced interpretation of Hilbert's early stance, which she argues does not place emphasis on anticipating "[Gödel's Completeness Theorem], nor on advocating a generalisable ontological principle; it is to demonstrate the advantages of his alternative and fruitful conception of consistency in mathematics." (Doherty [27], p. 117)

underlying presumption on which the debate turns fails.[24]

Tanswell [106] goes even further and argues that "mathematicians are a kind of conceptual engineers." By applying techniques and ideas from the field of conceptual engineering, we can—and should—play an active role in the development and role assignment of our concepts.[25] So foundationality can not only give *rise* to choice through open-texture, the property itself is also *subject* to change. Indeed, all the above-mentioned axes have been subject to change over time: the introduction of non-Euclidean geometries in the nineteenth century loosened pre-theoreticalness in geometry, and with the development of Robinson's non-standard analysis in the 1960s infinitesimals lost their ontological independence. Finally, proposals for multiversism could similarly influence how we think about the conceptual integration of the set concept.[26] This influence is, moreover, not just tracks the trajectory of mathematical interests, rather the assigned roles of set theory are actively debated within the community of practitioners. Tanswell's point is that such active deliberation should—and does—take place:

> I do not merely mean [the assertion that mathematicians are a kind of conceptual engineers] as a merely descriptive claim, but additionally as a methodology that should be embraced in mathematics. I demonstrated this as applied to the debate between advocates of the Universe view and the Multiverse view of set theory, arguing that the way forward in this debate is to understand the issue of tension as being of the right kind for conceptual engineering to resolve. Drawing on the resources of conceptual engineering will allow more powerful

[24]Note that Frege's idea of foundationality is very different from the Lakatosian one I have adopted. In fact, I have put Fregean concepts on the far *instrumental* side of the spectrum. Frege modeled his foundations after the Euclidean ideal, but as Lakatos writes:

> The story of Euclideanism is a story full of [...] dethronements. [...] The discovery of irrational numbers led the Greeks to abandon Pythagorean arithmetical for Euclidean geometrical intuition: arithmetic had to be translated into crystal-clear geometry. [...] The nineteenth century, in 'clarifying' the concept of irrational numbers, switched back to arithmetical intuition as the dominant intuition. Later on, Cantorian set-theoretical, Russellian logical, Hilbertian 'global', and Brouwerian 'constructivist' intuitions competed for this role. (Lakatos [62], p.12)

Morevoer, it should be noted that the mathematical milieu in which Frege and Hilbert debated differs substantially from the current context. Concepts from mathematical logic that are now standard, such as consistency, had not yet been settled in their modern form. See also Dean [26], pp. 349-51, and Doherty [27].

[25]For a general introduction to conceptual engineering see, e.g., Cappelen [22].

[26]About the set-theoretic multiverse program see e.g. Hamkins [42].

> insights into the creation and development of mathematical concepts. (Tanswell [106], p. 28)

Thus the foundationality of a concept, and so its Lakatosianess, is at least partially a matter of our current interpretation of mathematics. Geometry used to be foundational but has become instrumental, and this shift naturally triggered the Frege-Hilbert debate. On the other hand, category theory began as an instrument, but has since taken a more foundational role, for instance in Voevodsky's univalent foundations.

2.2 Conceptual Correctness

Now that we have a better grip on the concept/conception distinction and how it relates to open-texture, Lakatosianess, and foundationality, we are in the position to tackle the matter of conceptual correctness (Question 2). As we argued, we have a great deal of freedom in shaping our conceptions, but at the same time every conception must come with some notion of correctness. We will see that this poses a dilemma where we must either justify when a conception is admissible, or be able to provide direct explanations of conceptual correctness. Incurvati [54] has shown, however, that direct correctness arguments are already for set "marred with difficulties." The latter arm of the dilemma is therefore intractable, and so we will focus here on the problem of admissibility (Problem 2). It will become apparent that this problem turns on the permissibility of stipulation in conception formation (Problem 3), in particular with respect to circular justification. We will first introduce the Admissibility Problem (Subsection 2.2.1), then discuss why the demand for a notion of correctness gives rise to the dilemma and how this relates to concerns about stipulation (Subsection 2.2.2).

2.2.1 The Conception-Admissibility Problem

We can observe that the conceptual content which conceptions carry consists in practice of diagrams and descriptions that characterize the objects that the associated conception is about. An example of such a description in the literature is Boolos's account in [13] of the iterative conception of set:

> There are the null set and the set containing just the null set, sets of all those, sets of all *those*, sets of all *Those*, ... There are also sets of all THOSE. Let us now refer to these sets as "those." Then there are sets of those, sets of *those*, etc. (Boolos, op. cit., p. 8)

As a matter of terminology we will say that a description like this presents a *story* of a conception. For example, we may say that the story of the naive conception has it that the sets are predicate extensions, and that in this story "V is a set" is true. Likewise, when there is an intuitive model in the form of a diagram, we will speak of a *picture* of a conception. The cumulative hierarchy, for instance, we will say, provides a picture of the iterative conception.[27]

Boolos's story of set undoubtedly characterizes an admissible conception. However, what if we add to it, for instance, that the "etc." must be understood as signaling the continuation of iteration beyond the first measurable cardinal? The intuition here is clearly that this would not constitute an acceptable story of set, but it is not immediately obvious why that is precisely. It cannot just be a lack of faith in the consistency of measurable cardinals, as these are widely accepted among set theorists.[28] We thus run into what I will call the *Admissibility Problem*:

Problem 2 (The Conception-Admissibility Problem)**.** What are the limits of permissible description and depiction in the formation of conceptions in mathematics?

In the same article in which Boolos presents his iterative story of set, he writes that "our aim . . . is to analyze the conception [of set] we have, and not to formulate some imperfectly motivated conception that manages to imply the axioms."[29] This suggests that for Boolos the admissibility of a conception derives (in part) from the established use of that conception in practice. So, on this account, his story of set is admissible because it describes the sets as they are actually used by mathematicians, namely as sets of some things, with iterations of this idea along some small transfinite ordinal. A conception that arise in this way is also natural in the sense that "without prior knowledge or experience of sets, we can or do readily acquire it, easily understand it when it is explained to us, and find it plausible or at least conceivably true."[30]

These criteria of *actual-use* and *naturality* indeed get rid of some intuitively unacceptable stories, such as the above "etc."-amended description of the iterative conception. However, the naturality condition also rejects several conceptions (of set) that we do not want to disregard as inadmissible. The limitation of size conception, for example, arguably requires some knowledge of the set-theoretic

[27] Talk of stories and pictures here is, however, strictly a matter of convenient terminology. It should not be understood as indicative of any platonist or fictionalist leanings.

[28] Measurable cardinals are certain large cardinals that outsize the hyperinaccessibles, Mahlos, and weakly compacts.

[29] Boolos [13], p. 10.

[30] Boolos [13], p. 7.

paradoxes, so it fails the naturality condition.[31] But do we really want to say that this view is somehow illegitimate, to the extent that we should not even consider it? Likewise, the actual-use criterion leads to problems because its intention conflicts with the view which we defended above, namely that choice is pivotal in conception formation. Moreover, without an account of how conceptions may arise in practice, it is unclear what is even at stake, as of course any conception in use in some community is in 'actual use' in the straightforward sense. After all, mathematics is a distinctly human activity, so actual use arises with human social activity.

In the face of the Admissibility Problem, we may be tempted to say: "Well, the "etc."-amended story is not right, but this is not a matter of admissibility; any conception is *admissible*. The problem with it is rather that there is nothing about sets which should lead us to believe that the cumulative hierarchy extends the length of a measurable cardinal. So what that description says is simply not the case. It is incorrect." This provides a simple way out of the Admissibility Problem, but it also shifts all the responsibility for good conceptual hygiene onto the notion of conceptual correctness. I shall argue now that this yields a dilemma: either we have to resolve the Admissibility Problem, or we have to come up with a direct interpretation of conceptual correctness. However, as we will see, the latter arm of the dilemma has proven to be completely intractable, so we will not pursue this route here. Still, highlighting the dilemma clarifies the role of stipulation in conception formation. We will discuss this further in Section 2.3.

2.2.2 The Dilemma of Conceptual Correctness

Incurvati [53] has contested the strategy of showing, from general considerations about the nature of set, that the sets are contained in the cumulative hierarchy, noting that this approach is "marred with difficulties." The core of this strategy typically consists of an attempt to elucidate set by analyzing the structural properties of the presumed priority relation between sets, in which is grounded in the intuition that $\{\emptyset\}$ is 'prior' to $\{\{\emptyset\}\}$.[32] One approach to understanding this priority relation is to view it as priority of construction, suggesting that sets *really are* formed step-by-step through collecting together any available objects, perhaps by means of some idealized mental activity. This view runs into difficulties, however, because it struggles to explain uncountable and impredicative sets. An alternative is to conceptualize the priority relation as a form of metaphysical or ontological dependency. Incurvati discusses several ideas along these lines that have been pro-

[31] The limitation of size conception takes the sets to be those collections that are in some sense not 'too big'. We discuss it further in Chapter 3.

[32] The set-theoretic priority relation is discussed further in Section 3.2.

posed in the literature, but concludes that they all "seem to rely, at bottom, on some primitive notion of metaphysical dependence," and that while

> [such an] account based on a primitive notion of metaphysical dependence seems to be the only real candidate [we have] for an account which attempts to argue directly for the claim that all sets are those in the hierarchy [...] it is not clear to what extent we have a good grasp of the notion of metaphysical dependence, and, [that] even if we do [have such a sufficient grasp], there are major problems with showing that the claim that the iterative conception is correct about sets and levels follows from the account based on it. (Incurvati [54], pp. 75-9)

It is therefore doubtful whether we can come up with a direct interpretation of conceptual correctness that not only covers the notion of set, but that can be applied to conceptions generally. Incurvati proposes instead to correlate correctness with reasons to believe that "[the iterative conception] is better than its rivals with regard to certain virtues that a conception of set may have."[33] This "indirect strategy" of *inference to the best conception* offers a tractable account of correctness that can draw for support on the existing literature surrounding the analogous principle of inference to the best explanation (IBE) in the philosophy of science. Moreover, it does not commit us to such controversial notions as analytic or conceptual truth.

Boolos [12], [13] likewise supports an indirect approach to correctness, and emphasizes herein the justificatory role of conceptions (Question 3):

> The iterative conception is the only natural and (apparently) consistent conception of set we have, and it implies [$\mathbf{Z}-$ Extensionality]; *that* is the justification (if that is the right word) it provides for [$\mathbf{Z}-$ Extensionality]. (Boolos [13], p. 7, emphasis as in the original.)

Incurvati ([54], p. 23) explains that the standard strategy to show that the axioms of a theory T are justified on a certain conception C is to "appeal to an intuitive reading of the axioms of T, to what they express, and say that what they express is the case on that conception." But, suppose that we do not restrict what constitutes an admissible conception, then explaining the "etc." in Boolos's description as suggested above, yields a legitimate conception of set that supports a theory much stronger than $\mathbf{Z}-$ Extensionality. On this conception it is clearly the case that there is a measurable cardinal. Accepting Incurvati's principle, this conception

[33] Incurvati [53], p. 84.

must therefore be preferred over the original as (more) correct.[34] The upshot is that our *stipulation*—that the cumulative hierarchy extends beyond the first measurable cardinal—can immediately be cashed out as a true property of the sets, thereby justifying the existence of a measurable *tout court*. This reasoning is surely viciously circular. And here we have our dilemma, one of the two assumptions must go. But as we argued in the section on open-texture (Section 2.1.1), without any alternative criteria to judge between conceptions, if we take the latter horn of the dilemma, the notion of conception itself becomes obscured. Therefore, we will not pursue this approach.

Still, isolating the dilemma here shows that the Admissibility Problem can be recast as a matter of framing the role of stipulation in conception formation. What form this role should take is by no means clear-cut, since in some cases, such as with the axiom of infinity, the same apparent circularity is not regarded with similar suspicion. Therefore, careful control must be exercised over how stipulation may feature in legitimate pictures and stories of conceptions, or it will threaten what conceptions are considered to *do*, namely to sharpen a concept and not only act as a mere formalism.[35] Hence, we will focus in the upcoming section on the *Stipulation Problem*:

Problem 3 (The Stipulation Problem)**.** What is the role of stipulation in conception formation with respect to the kind of circular justification highlighted here?

In Subsection 2.4.4 we return to the example of justifying Infinity versus measurable cardinals.

2.3 Stipulation, Freedom, and Coherence

What is so bad about stipulation? Do we not stipulate, for instance, in hyperbolic geometry that for any given line ℓ and point P there are distinct lines parallel to ℓ through P? And do we not also stipulate that groups have an inverse for each element under the group operation? Waxman [112] argues that throughout the history of mathematics, there has been a prevailing inclination towards liberalism, allowing mathematicians to explore concepts and axioms that they find interesting or useful. He proposes that this supports the following thesis:

[34]This assumes, of course, that one of the possible virtues that a conception can have over another, is that it supports a stronger theory, but this seems inevitable on the indirect strategy.

[35]Incurvati ([54], pp. 14-20) identifies sharpening as what "a philosopher (or a mathematician, or a logician) [does] when she puts forward a conception of set." He identifies two more purposes of conceptions, namely analysis and replacing inconsistent conceptions, but the arguments presented here can be easily adapted to these goals as well.

> **Epistemic Freedom**: Mathematicians can *justifiably* work within and draw out the consequences of any theory meeting minimal standards of logical *coherence*, in the sense that doing so manifests no epistemic error. (Waxman [112], p. 6, first emphasis is in the original, second is mine.)

The constraints on stipulation in conceptualization should therefore be derived entirely from considerations of *coherence*, or as Boolos puts it, from the question of whether there is "a thought behind [the stipulations]."[36] We now consider two possible interpretations of this sense of coherence. One option we already discussed in Subsection 2.1.2, namely the liberal hilbertian stance that consistency suffices for existence, and thereby for coherence. Another more recent suggestion due to Isaacson [55] interprets coherence as a matter of *use in practice*. I will argue, however, that both proposals fail to appropriately tame what can be stipulated. In the following section I then plead that coherency may be derived instead from explanation.

2.3.1 The Hilbertian Criterion

The hilbertian understanding of existence, and *a fortiori* of coherence, as depended only on consistency, is historically influential, but it fails as a criterion of permissible stipulation in our sense.[37] As a result of Gödel's Second Incompleteness Theorem, there is no computably enumerable background theory of arithmetic that we can adopt which fixes the question of consistency for all theories. Suppose that we are led to believe (perhaps inductively) that some theory T is consistent, and suppose furthermore that we have taken **PA** as our arithmetical reference point for the hilbertian criterion, and that T interprets **PA** (i.e. is not too weak). By Gödel's theorem, which can be proved in **PA**, both $\mathbf{PA} + \mathrm{Con}(T)$ and $\mathbf{PA} + \neg\mathrm{Con}(T)$ are consistent theories. But the criterion then also applies to these two theories, and thus each of them will be a coherent *theory of arithmetic*, while disagreeing about the consistency of T. This relativity casts doubt on the absoluteness of the notion of consistency itself, and therefore on the meaningfulness of the "consistency implies coherence" criterion. Also, to move the reference theory from **PA** to $\mathbf{PA}+\mathrm{Con}(T)$ on the basis of some other argument or conviction (perhaps inductive), is essentially to remove the antecedent of the criterion. But

[36] In any case, this is my reading of Boolos [12], p. 219.

[37] Surely the existence of the objects described by some conception implies the coherence of that conception. Also, Hilbert's own (early) views are debatable, hence I have chosen to refer to the criterion here as broadly hilbertian (without capital). For one perspective, see Doherty [27].

then the criterion no longer does any work, as we could just as well have taken that argument or conviction directly as the justification of coherency.[38] Moreover, a base theory which, unlike **PA**, can settle the consistency of at least the theories T that are of practical interest, is not forthcoming, since it would have to be able to carry out arguments that cannot be formalized in our best theory of set. And at any rate, consistency in general is a hard problem. The set of Gödel numbers of consistent first-order sentences is Π_1-complete, and many intractable problems in number theory, like the twin prime conjecture and the Riemann hypothesis, fall in this class.

2.3.2 The "Use in Practice" Criterion

A more promising route to coherency has recently been articulated by Isaacson [55]. Building on Kreisel's notion of *informal rigor*, he argues that mathematics is at its core conceptual, and that we are justified in assuming the existence of a particular structure provided we *show* that it is coherent. This proceeds by "developing our mathematical understanding of the subject matter of [a] theory through informal rigor," which in turn is achieved not by "axiomatizing the notion, but by reflecting on the development of mathematical practice by which particular mathematical structures come to be understood, the natural numbers, the Euclidean plain, the real numbers, etc."[39]

Isaacson makes a distinction here between *particular* and *general* mathematical structures to aid his structuralist account of the sets, which is an especially hard case for structuralist philosophies of mathematics.[40] He writes:

> Mathematicians study two sorts of structures, which I shall call particular structures and general structures. The distinction is marked by use of the definite and indefinite articles. We speak of the natural numbers and a group. Particular structures include the natural numbers, the Euclidean plane, the real numbers. General structures include groups, rings, fields, metric spaces, topologies. The particularity of a

[38] There are also mathematical issues with too much existential tolerance. As a result of Gentzen's consistency result that Con(**PA**) can be proved in **PRA** plus quantifier-free induction up to the ϵ_0, there are models of **PA** where (a suitably coded version of) quantifier-free induction up to ϵ_0 fails. Thus relativity about the natural numbers leads to relativity about the ordinals, and therefore relativity also about sets. This example exhibits the conceptual integration axis of the foundationality of the natural numbers (Subsection 2.1.2).

[39] Isaacson [55], p. 29.

[40] According to Isaacson: "Establishing an account of set theory is a key test for any would be structuralist philosophy of mathematics." ([55], p. 3)

> particular structure consists in the fact that all its exemplars are isomorphic to each other. The generality of a general structure consists in the fact that its various exemplars need not be, and in general are not, isomorphic to each other. (Isaacson [55], p. 21)

This distinction suggests the following answer to the Stipulation Problem: stipulation is free for general structures, since they are tied to instrumental concepts. On the other hand, particular structures, which arise from foundational concepts, require stipulations to be supported by a conception about which we have become clear through using it in practice (i.e. informal rigor).

Still, this alone will not do as Hamkins [41] observes:

> I claim that it is precisely this kind of mathematical experience that set-theorists have gained with respect to the various (particular) forcing extensions of the universe and the other models of set theory commonly studied. We know what it is like to live in the universe obtained by adding $\aleph_2$ many Cohen reals over L and what it is like in the $L(\mathbb{R})$ of that extension. We know what it is like in the universe obtained by forcing MA + $\neg$CH over L, or by forcing with Sacks forcing [...] We have sufficient experience living in these diverse set-theoretic worlds to know that the concept of set used in each of them is perfectly robust and satisfactory as a concept of set. Each of these universes feels fully set-theoretic. (Hamkins [41], p. 139)

The essence of the argument here is that set theorists, including Hamkins, possess the kind of informal rigor experience that Isaacson claims substantiates the coherence of a particular structure, not just for a single true story of set, but for many mutually incompatible stories, generated, e.g., by forcing. This situation is unacceptable for Isaacson, however, because he believes in the applicability of Zermelo's categoricity result, which already settles CH, for instance. Therefore, his account of coherence is lacking, even on its own terms.

However, the practical experience of Hamkins does not necessarily refute informal rigor as a coherence criterion for stipulation on our account of admissible conceptions, which does not accord the importance that Isaacson attaches to categoricality results such as Zermelo's. Still, Hamkins's goal here is at cross-purposes with our own, as <u>set</u> according to him, although ontologically independent, is not really foundational. In fact, he explicitly draws a comparison with the situation of post-nineteenth century geometry:

> At first, these alternative geometries were considered as curiosities [...] In time, however, geometers gained experience in the alternative ge-

> ometries, developing intuitions about what it is like to live in them, and gradually they accepted the alternatives as geometrically meaningful. Today, geometers have a deep understanding of the alternative geometries, which are regarded as fully real and geometrical. The situation with set theory is the same. The initial concept of set put forth by Cantor [...] seemed to be about a unique concept of set [...] The multiverse view now makes the same step in set theory that geometers ultimately made long ago, namely, to accept the alternative worlds as fully real. (Hamkins [42], p. 425-6)

The type of coherence implicit in Hamkins [41] may perhaps suffice to confer ontological independence—and this *is* his primary concern as a multiverse adherent. However, the foundational tasks that we have entrusted set theory with also require a correctness criterion which, as we have argued, should in principle allow us to settle open-texture in a reasoned manner. Therefore, the question of *why* something is the case on a conception relative to some alternatives should, at least, be meaningful. With Hamkins's forcing-generated conceptions of set, there seems to be no rationale in asking why, for instance, CH is true (or false) rather than false (or true) on any of these conceptions, since its truth value is entirely stipulative. There is therefore no room for explanation, and this is not a matter of the tractability of the question.[41]

Thus, being able to ask and answer why-questions is *necessary* for establishing coherence, and Isaacson's "use in practice" criterion fails to respect this demand. In the next section I will argue that explanation can also act as a *sufficient* criterion for coherence.

2.4 Explanatory Coherence and Why-Questions

2.4.1 The Role of Explanatory Saturation in Coherence Arguments

Our insistence on coherence in foundational concepts is essentially a demand for stability within our mathematical conceptual framework. Coherence justifies our assumption that we can do meaningful work within this framework, as it is supposed to shield us (to some extent) from the following potential conceptual problems:

- **Inconsistency**: incongruent principles supported by a conception;

[41] A further concern that I have with Hamkins's universes as admissible conceptions of set is that they can only be formulated relative to a prior conception of set.

- **Mathematico-Philosophical Incongruence**: conflict with the mathematico-philosophical background beliefs.
- **Improper Conceptualization**: failure to meet pre-theoretical constraints;
- **Integration Failure**: failure to respect conceptual dependencies;

The intuition that explanation can aid in averting the first two types of conceptual problems—which both turn on some form of inconsistency—is quite evident. The set-theorist Steel writes, for instance, that: "A voluble witness with an inconsistent story is more likely to contradict himself than a reticent one."[42] Similarly, Incurvati contends: "Being able to explain the paradoxes can be taken as such a criterion [of conceptual satisfactoriness] *because* consistency in a theory is a virtue."[43] Furthermore, improper conceptualization of an idea constitutes a direct explanatory error since the pre-theoretical constraints form an integral part of explaining what the idea entails.

As for conceptual integration, a historical example of the role that explanation can play herein is found in Cantor's justification of his transfinite numbers (the ordinals). Cantor showed how the transfinite numbers could be conceptually integrated with the finite numbers by extrapolating certain basic (arithmetical) properties.[44] For this, he drew inspiration from his concept of set and how it can be related to number via the notion of well-orderedness. The interrelation at play here, he derived in turn from a general methodological heuristic for conceptual integration of the finite with the infinite. This "Cantorian finitism" ultimately forms the core of Cantor's legitimization of embracing a certain structural similarity between the finite and the completed infinite (transfinite).[45] Hallett encapsulates this integration argument as follows:

> [...] since [the natural numbers] are enumerals of finite well-ordered sets, *and* since the concept of well-ordered set and enumeral is framed quite generally, surely transfinite enumerals are legitimate (exist) too. [...] This would be, then, a kind of argument by analogy, analogy

[42] Steel [101], p. 156.

[43] Incurvati [53], p. 85, emphasis is mine.

[44] This sort of integration is called "unification" in Michael Hallett's standard work on Cantorian set theory [40].

[45] Cantor's inspiration for this 'finitism' was primarily theological (see Hallett [40], pp. 32-40, and in particular Cantor's quotation of Augustine on pp. 34-5). Besides this controversial intrinsic justification, however, the very possibility of the unexpected unification that is suggested by this Cantorian rule of thumb surely also provides some circumstantial evidence for its validity. Maddy ([74], p. 486) mentions another extrinsic justification.

> supported by Cantor's principle of 'finitism'. The concept of transfinite enumeral extends a perfectly coherent basic concept, and since there are finite enumerals (numbers) corresponding to finite sets, why not also transfinite enumerals (numbers) corresponding to the infinite well-ordered sets? (Hallett [40], p. 53; emphasis as in the original.)

Another historical example that highlights the significance of explanation in bolstering or challenging coherence is Berkeley's famous critique of the Newtonian and Leibnizian calculus in *The Analyst* (1734). Berkeley's complaint was essentially an explanatory challenge to the analysts to clarify how it is that infinitesimal magnitudes can be discarded whenever convenient, and yet that these are manifestly not zero. He protested: "And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?" According to Berkeley, the inability of the analysts to satisfactorily address these questions cast scientific doubt on the entire framework of calculus, despite its successful applications in geometry and apparent consistency.[46] Boman [11] relates:

> Calculus, according to Berkeley, was a house built on sand. A minor shift of the sand would bring the whole structure tumbling down. However, the problem is, [as even Berkeley agreed], that [...] calculus works! [...] [Against this, Berkeley held] that "the modern analysts" got correct results from false assumptions because they made two errors that compensated for each other. (Boman [11], p. 64)

These considerations, combined with those explored in the previous section, suggest that the coherence we are striving for is explanatory in nature, or at the very least, they imply that coherence can be attained through explanation. In other words, my contention is that when we are engaged as conceptual engineers with specifying a foundational concept, it suffices to be able to justify the stipulative conceptual choices made by explaining them in terms of other conceptual choices or in relation to our mathematico-philosophical background beliefs.[47] The threshold of appropriate explanation here should be determined by the community

[46] Berkeley's point was to argue that the "infidel" analysts were guilty of the same kind of unscientific leaps of faith (i.e. not up to the Euclidean ideal) which they held made the Christian religion incomprehensible. See Sherry [99] and Boman [11]. Berkeley's quote about the "ghosts of departed quantities" is taken from Boman.

[47] This demand is resembled by a criterion of Boolos (as identified by Paseau [86], p. 37) that "The obviousness of [a principle] P must not be independent from that of the rest of [the conception] C."

of practitioners. Moreover, given the crucial role that we attribute to explication in conceptualization, *explanatoriness must be considered as one of those virtues of Incurvati against which the correctness of conceptions can be compared.* So, the greater the number and quality of explanations, the better. Hence, we should strive for *explanatory saturation*: everything that can be explained should indeed be explained, and with the greatest power.

An analogy with fiction might be helpful here: a good (admissible) story has an explanatory richness which allows and guides the development of the story beyond the original narrative. The stage for the Greek hero Odysseus, for example, is set in the Iliad and the Odyssey, which serve as a kind of founding context (similar to a foundational concept). This context guides and constrains (provides a sense of correctness for) the development (sharpening) of the story of Odysseus. So, although the tale (the stipulation) of how Odysseus tried to feign madness by sowing his fields with salt postdates the Iliad and the Odyssey by many centuries, it nevertheless could become part of the modern canon (conception). The episode harmonizes (can be explained) well with what was already understood about the cunning Odysseus (the concept), and it therefore provides a coherent account of some aspects that interest us but were left open by Homer (open-texture).[48] On the other hand, stories without such an explanatory cadre (like instrumental concepts) cannot be extended in a non-arbitrary manner, because then there would be no frame of reference that could possibly ground the extension. You cannot issue an opinion on whether the bishop in chess should only move diagonally, and hence it is not subject to open-texture.

2.4.2 Explanatory Coherence in Mathematics

The idea of inter-connected explanations as a source of coherence ties back to an old debate in epistemology between *coherentist* and *foundationalist* theories of epistemic justification. Foundationalism in epistemology paints a Euclidean picture of justified belief as something that 'flows' from self-evident truths to complex beliefs

[48]Similarly, the foundational narratives of humanity, as found in texts like the Torah, the Qur'an, or even the U.S. Constitution, set up broad systems of thought that are subject to extensive exegesis and supplementation, through, e.g., Talmudic commentary, hadith (a loose body of records of events that involve prophet Muhammad), Amendments, etc. However, despite being open to interpretation, these analyses of and supplements to foundational narratives are decidedly not arbitrary or incoherent, rather they (attempt to) harmoniously complement what went before. When the Supreme Court or a religious authority issues an opinion, it is typically accompanied by a substantive piece of writing that summarizes the facts and case law and addresses the legal or religious concerns that arise from it.

by means of logic.[49] Coherentism replaces this top-down justificatory structure in favor of a holistic "web of beliefs," which has no favored basic tenets. Justification is instead derived from the mutual support provided by this web. The idea is that the harmonization so effectuated permeates the entire system with justification, even if the individual beliefs, when taken in isolation, may not be justified, or at least not to the same extent. When such mutual support takes the form of intertwined explanations, we obtain what is called *explanatory coherence*.[50]

The nature and applicability of explanatory coherence is a controversial matter that is embedded in a very substantial literature. Hence, I will not be able to settle the matter here.[51] However, it is important to keep in mind that, because we are not after empirical truth, but only after admissibility, the function we need explanatory coherence to fulfill is significantly less demanding than what it is supposed to achieve when applied to science. We only need it to help us justify certain stipulations in the face of the stability issues mentioned at the beginning of Section 2.4. This allows the usual (general) arguments against coherentism to be addressed rather straightforwardly: according to *the isolation objection*, coherentism fails because facts about the external world are not interconnected with other beliefs but are directly grounded in experience, making it impossible to integrate them into the web of beliefs. However, since we are interested in the coherence of abstract things, which do not inhabit the world around us, this poses no particular problem. Likewise, *the alternative coherent systems argument* presents no issue, since incompatible coherent systems do not clash when our aim is not truth but only admissibility. Conceptual correctness—which is still not the same thing as conceptual truth—comes only after we have determined admissibility, and not by means of coherence, but through inference to the best conception. For this same reason, we also do not run into any *feasibility problems* of having to check for consistency to establish coherence, as the use we put our notion of explanatory coherence to does not necessitate consistency.[52] It would be strange indeed to think

[49] As we discussed in Subsection 2.1.1, there are strong reasons to reject such foundationalism in mathematics. See, e.g., Lakatos [63], [62]. The metaphor of the flow of truth originates with Lakatos.

[50] The idea of coherence through explanation can be found in, e.g., Thagard [108], who applies the idea to science.

[51] For an overview of several arguments against explanatory coherentism and some attempted rebuttals of these, see Lycan [73]. A book-length defense can be found in Poston [88].

[52] This argument also blocks *the preface paradox*: suppose that a very careful academic publishes a well-researched book on history. Because of her extensive investigation and the cautious editorial process, she has justified belief that each of the claims she defends in the work is in fact true. Yet, at the same time, she honestly admits (in the preface) that no doubt some errors have crept in her book, since that is bound to happen in history research. That is, she justifiably be-

that the naive conception of set was always incoherent, or even inadmissible, since it was widely adopted at one point among practitioners. When inconsistencies do arise in foundational concepts, they just present further explanatory challenges that prompt conceptual sharpening or revision of the mathematico-philosophical background believes. There is nothing really explanatorily incoherent about a conception in which a contradiction may arise but has not yet been discovered. Moreover, we do not even necessarily have to explain away every inconsistency. An explanation of some inconsistency may also involve the background logic. Priest's dialetheism does this, for example, by providing a general philosophical framework for dealing with inconsistencies.[53] [54]

2.4.3 Why-Questions

We have not yet addressed the mode by which explanation is to be effectuated. Pease et al. [87] have applied empirical techniques from experimental philosophy on data gathered from online collaborative mathematics projects to argue that explanations in mathematics are answers to *why-questions*. They looked for keywords that signal explanation, such as "expl*" and "underst*", and then analyzed the context to assess whether explanation was indeed taking place, and if so, whether it could be construed as answering a why-question. Their results showed that 37% of the comments in the corpus contained some form of explanation and that 99% of these explanations could be read as answers to (implicit) why-questions.

These findings run counter to the arguments put forth by Zelcer [114] and others against the possibility of mathematical explanation. Pease et al. ([87], p. 17) propose that this discrepancy may originate from the fact that explanation has traditionally been analyzed at the "frontstage" of mathematics (published proofs), while their focus is on the "backstage" (informal mathematical discussion).[55] Considering that mathematics in progress primarily involves the latter, I take it that explanation in relation to conceptualization should be understood from a backstage

lieves that at least one of the claims she makes is actually false, yet at the same time she believes, also justifiably, that each of her claims is true. In other words, she is justified in believing an inconsistent set of claims. But then, as logical inconsistency is an obvious form of incoherence, how can coherence be necessary for justification? Or so the argument goes. However, because we do not assume that coherence implies consistency, this paradox poses no particular issue for us.

[53] See Priest's [92] for a book-length philosophical introduction to dialetheism and paraconsistent logics.

[54] The list of common arguments against coherentism, which we have all countered here, comes from the IEP entry on "Coherentism in Epistemology"([81]).

[55] The front/back distinction originates with Hersh [47].

perspective, and that therefore the arguments of Pease et al. apply.[56]

Finally, I do not mean to suggest that why-questions are the *only* possible avenue of explanation. It is enough here to argue that why-questions form an integral part of explanatory coherence. But in the next chapter (Section 4.3) we will also come across a what-question that is intimately tied up to a why-question: why is V not a set, and what is it instead?

2.4.4 The Justification of Infinity and Measurable Cardinals

Let us now return to the problem I sketched in Section 2.2 about justifying Infinity versus justifying measurable cardinals. Why are there infinitely many sets? This question can be answered, for example, by appeal to the imaginative capacity of visualizing a model—as recently attempted by Waxman [111]—or by a general Cantorian believe that every potential infinite presumes an actual infinite:

> One can also demonstrate incontestably the occurrence of the actual infinite [...] for since there can be no doubt but what we cannot do without variable magnitudes in the sense of the potential infinite, then the necessity for the actual infinite can be proved as follows: In order for such variable magnitudes to be capable of evaluation in a mathematical investigation, the "range" of their variation must be precisely known by means of a prior definition; but this "range" cannot itself be in turn something variable, for otherwise every fixed support for the investigation would give way; hence, this "range" is a definite actually infinite set of values.
>
> *Thus every potential infinity, if it is to be employable rigorously in mathematics, presupposes an actual infinity.* (Cantor 1886, quoted in Hart [44], p. 249, emphasis is mine.)

The point is that these responses appeal to, or engage with, our overarching mathematico-philosophical framework, which every conception has in common. Indeed, if this background is finitistic, for instance, we have to acknowledge even Boolos's original story of set as inadmissible, since the stipulative "..." would then be incoherent.

---

[56] There is a resemblant account of explanation in the philosophy of science due to van Fraasen [35], who introduced the idea of contrastive why-questions. It would be interesting to see whether a uniform interpretation of explanation across mathematics and the natural sciences can be achieved along these lines. This would also fit well with Lakatos's quasi-empiricist view of mathematics.

On the contrary, the existence of measurable cardinals does not follow in the same manner because, "unlike the small large cardinals suggested by *inexhaustibility*, measurable cardinals are not usually held to follow naturally from the concept of set or the nature of the iterative hierarchy."[57] Thus far, there has been no satisfactory explanation as to why these cardinals should exist, neither in terms of anything inherent to the story of the iterative conception of set, nor with respect to the general mathematico-philosophical context. Efforts to justify the existence of measurable cardinals based on their utility or apparent consistency do not align with our intended understanding of explanation. This also holds for attempts to defend large cardinals on the basis of structural connections to axioms of determinacy. Koellner phrases this line of thought as follows:

> The hope is that a series of deep mathematical theorems will reveal structural connections [between large cardinals and axioms of determinacy] that provides more and more evidential support for the statements. Should this be the case then the loss in evidence attached to the minimal points [i.e. the relatively trusted axioms of **ZFC**] as one climbs the interpretability hierarchy would be compensated for by the gain in structural connections that are revealed as one ascends to richer domains of mathematics. (Koellner [60])

While these structural connections are interesting in their own right and may lend support to large cardinals, the evidence here is tied to a theory, and not to a conception. It is therefore not admissible as explanation in establishing conceptual coherence.

One may object that this evidential distinction is contentious because it parallels the tenuous difference between extrinsic and intrinsic justification (Subsection 1.1.1). For example, Barton et al. [3] have recently challenged the tractability and demarcation of these justificatory categories. Hence, we should address their criticism to maintain our response to Koellner.

2.4.5 The Tractability of Conceptual Justification and its Demarcation from Extrinsic Evidence

With regard to the tractability of intrinsic (conceptual) justification, Barton et al. identify two main issues:

> [...] one [issue] is the absence of a clear methodology through which one can ascertain that a statement is 'derived' from the concept of set;

[57] Maddy [74], p. 505, emphasis as in the original.

> the second one is the fact that there are many conceptions of set, all of which may give rise to different versions of 'intrinsicness', and, thus justify some axioms rather than others on intrinsic grounds. (Barton et al. [3], p. 167)

The second issue raised here does not surface on our account of conceptualization in mathematics because justification for some axiom X on a conception is not the same thing as an argument for the correctness of X *tout court*.[58] However, the first point of criticism is well-taken; while answering why-questions serves as a *method* for demonstrating justification, this approach is still predominantly a matter of contentious philosophical debate rather than subject to clear mathematical proof. On the contrary, Barton et al. ([3], p. 181.) provide a "fully technically precise notion" of justification as explanation rooted in *prediction* and *confirmation*: our confidence in an axiom Ψ increases if $\mathbf{ZFC} + \Psi$ predicts some result ϕ that was previously unknown and later confirmed to follow from **ZFC** alone. But, as I will argue now, this approach fares no better than ours in avoiding debate and thus is equally intractable.

There are numerous examples of prediction and confirmation where the justification conferred seems debatable: Keisler [56] proved the Keisler-Shelah isomorphism theorem in 1961 using the generalized continuum hypothesis (GCH). Then, in 1971, Shelah [98] gave a proof that avoided GCH altogether. But how should this confirmation of GCH weigh against Woodin's set-generic multiverse program, which implies ¬CH?[59] This is surely a matter of debate. Also, Baumgartner and Hajnal [5] proved the partition relation $\omega_1 \to (\alpha)^2_2$ first from Martin's Axiom (MA), then used an absoluteness argument to lift the result from a forcing extension where MA holds back to V, thus providing confirmation for MA. Furthermore, we have many results which show that "in some sense even the tiniest strengthening of MA fixes the continuum at $\aleph_2$," so this likewise suggests ¬CH.[60] Again, the conflicting implications here can only be settled by (philosophical) discussion.

This leaves the alleged demarcation problem:

> [...] the force of an intrinsic justification cannot be neatly separated from that of an extrinsic one. In the set-theoretic context even [the] cases where one might have the impression that such a demarcation

[58]This preempts some Quinean worries that might arise from any correspondences between intrinsicness and analyticity.

[59]For a brief discussion of Woodin's set-generic multiverse program and Ω-logic in context, see Ternullo [107], pp. 57-9. An extended discussion is found in Koellner [61].

[60]The quote is from Magidor [78], p. 18.

> has been clearly drawn are, on closer inspection, difficult to demarcate. Whilst there is a sharp distinction between a principle itself and its consequences, this is not so for the justificatory force attaching to intrinsic and extrinsic justifications. (Barton et al. [3], p. 170)

For example, Gödel held that maximizing the 'richness' of the power-set operation harmonizes well with the (iterative) conception of set. Hence, he (later) dismissed $V = L$ on the ground that it limits the plenitude of the subsets. Thereafter, consequences of the axiom of constructibility were discovered that bolstered this intrinsic justification, most notably Scott's result that $V = L$ implies that there are no measurable cardinals. These extrinsic considerations reinforced the intrinsic belief that $V = L$ is restrictive, while this belief in turn had suggested that there must be such results. Hence, there may be mutual interaction between extrinsic and intrinsic considerations, and therefore they cannot be untangled and sharply demarcated, or so the argument goes.

Regardless of the persuasiveness of this example, however, it does not pose any particular problem for our account of conceptualization in mathematics. If we have a conception of set that supports say **ZFC** + $\exists\mu(\mu$ is measurable), we may incorporate Scott's theorem in the answer to the question why the powerset operation creates non-constructible sets. Well, it does so because we have reasons X, Y, and Z to justify **ZFC** and the existence of measurable cardinals on our conception of set, and Scott's result then shows (by logic) that there is a non-constructible set, that is $V \neq L$. This seems to be as good an explanation as any, since we are not pre-theoretically inclined to belief that $V = L$. Moreover, from the perspective of a conceptual engineer, which we have adopted, there is no error in making conceptual changes to incorporate extrinsic evidence. The demarcation problem is therefore not particularly acute on our view.

2.5 Conclusion

I have developed a Lakatosian account of conceptualization in mathematics which emphasizes the role of the community of practitioners. This account explains why there are conceptions of set, but not conceptions of group, and how conceptual correctness, open-texture, and concept sharpening are intertwined. We saw that the considerable epistemic freedom we have in shaping our conceptions poses a dilemma that brought the Admissibility Problem (formulated as Problem 2). This was then reframed as primarily a concern about circular justification of stipulations (Problem 3). I subsequently proposed to tie the permissibility of stipulation in a story or picture of a conception to the explanatory coherence of the overall system

of conceptions and the mathematico-philosophical context. I argued that such coherence can be effectuated by answering as many why-questions as possible, and to our best ability, and that this simultaneously demonstrates a virtue that can be used as a yardstick to compare the correctness of conceptions. Finally, I applied this account of conceptualization to argue that the stipulation of Infinity in the iterative conception is warranted in a sense that does not transfer to the stipulation of measurable cardinals.

There are of course also some problems with the presented framework. The most glaring issue concerns the tractability of explanatory coherence. How many and which why-questions should we be able to answer about a conception before we can deem it to be coherent? The cheap way out is to say that alternative accounts do not fare better in this regard, and I indeed argued in this spirit in Subsection 2.4.5 against Barton et al. [3]. However, I am inclined to believe that if we adopt the perspective of Lakatos and Tanswell on the role of the community of practitioners in the process of conceptualization, it is unlikely that we can ever formulate such precise criteria. Another concern is that more targeted anti-coherentist arguments can be formulated within the domain of mathematics, which cannot be as easily defused as the standard arguments we discussed in Subsection 2.4.2.

Despite these drawbacks, I hope to have stated a convincing case for the significance of why-questions for conceptualization in mathematics. In Chapter 4, I argue against Soysal [100] that Linnebo's potentialism about the iterative conception of set provides genuine explanatory power in addressing why V is not a set. Consequently, by harnessing the import of why-questions which I have established here, I conclude that potentialism should be preferred over minimalist approaches like Soysal's. But first, as a warm up, in the next chapter I present an argument along the same lines to show that the iterative conception is preferable to the limitation of size conception, and therefore should be considered (more) correct. This chapter also functions as an introduction to and discussion of Linnebo's potentialistic modal plural set theory [69] [68].

Chapter 3
Why is V not a Set?

It is time to put the account of conceptualization we have developed in the previous chapter into practice through a case study. Historically, the set-theoretic paradoxes (listed in Subsection 1.2.1) gave impetus to a variety of developments of the early concept of set. We have already come across the iterative conception, or sets as "set of . . . ", and Cantor's consistent/inconsistent multiplicities (Subsection 1.2.1). The latter has been understood as a form of *limitation of size*, where the sets are taken to be those collections that are not 'too big', in the sense that V does not inject into them.[1] The purpose of this chapter is to argue that, based on the significance of answering why-questions, the iterative conception is preferable to the limitation of size conception with regard to explaining the set-theoretic paradoxes, and therefore should be considered more correct.[2] This chapter also serves to introduce and discuss Linnebo's potentialist account of the iterative conception.

The plan is as follows: against the background of the potential conceptual problems set out in Subsection 2.4.1, we propose some criteria in the next section by which to judge explanations of the paradoxes. We argue that Boolos's naturality condition (which we already mentioned in Subsection 2.2.1) preempts a viciously circular answer to the question why, for instance, V is not a set. The limitation of size conception therefore does not provide a coherent explanation of the paradoxes, since Boolos argues that it is not natural in this sense. Then, in Section 3.2, we introduce Linnebo's potentialist account of the iterative conception of set [69] [68] and argue that this *does* accord with the criteria we have set up. In the next chapter we compare this account of the iterative conception against the minimalist account of Soysal. There I will present a novel argument that potentialism is not merely an "idle wheel," which Soysal argues, because it does not run into the explanatory instability that I will uncover in her minimalist account.[3]

[1] Hallett [40] is the *locus classicus* of the limitation of size interpretation of Cantor's work. See also Incurvati [54], chapter 5.

[2] Recall that we argued in the previous chapter (Subsection 2.4.1) that explanatory power is also a virtue by which correctness should be assessed.

[3] Soysal [100], p. 591.

3.1 What Makes a Good Explanation of the Paradoxes?

Whether a conception of set can be regarded to provide an acceptable account of the set-theoretic paradoxes, should be judged on its ability to shield against inconsistency and mathematico-philosophical incongruence (Subsection 2.4.1) in the face of the uncertainties introduced by these paradoxes. We postulate some explanatory virtues that such an account should have in this sense:

- ➢ NATURALITY: The conception of set used to explain the paradoxes should not itself directly incorporate our knowledge of the paradoxes. That is, it should not be ad-hoc.[4]

- ➢ DEPTH: An account of the paradoxes offered by a conception of set should not merely identify how the conception blocks the paradoxes on a superficial level, such as with respect to syntax.[5] It should provide a deep explanation that does at least the following:

  - identify all the conflicting pre-theoretical assumptions which are relevant to the proposed resolution of the paradoxes;
  - conceptually clarify how these conflicting assumptions combine to yield contradiction, and argue, or at least make plausible, which of them is the culprit on the propounded conception of set.
  - unambiguously force an answer to how this culprit must be neutralized to remove the known inconsistencies.

- ➢ SIMPLICITY & UNIFORMITY: Explanations should not involve elements that serve no or a redundant explanatory purpose (no idle wheels); they must be as simple as possible. Moreover, ceteris paribus, they should explain the paradoxes uniformly to avoid overfitting and ad hocism.[6]

- ➢ (MATHEMATICAL RICHNESS: An explanation should not depend on a mathematico-philosophical framework that demands impracticable modifications to the

[4]See for instance, Boolos [13], p. 7.

[5]Boolos ([12], p. 219) puts it as follows: "A final and satisfying resolution to the set-theoretical paradoxes cannot be embodied in a theory that blocks their derivation by artificial technical restrictions on the set of axioms that are imposed only because paradox would otherwise ensue." (emphasis as in the original.)

[6]Forster ([33], p. 9) puts it thus: "A single explanation for a lot of hitherto apparently unconnected phenomena is prima facie more attractive than lots of separate explanations. Unifying-single-explanation arguments are common, natural, and legitimate [...]"

> mathematical practice as it exists. It should not impinge on the subject's progress.)[7]

Boolos has argued that the limitation of size conception does not possess NATURALITY:

> Unlike the naive and the iterative conceptions, limitation of size [...] is not a *natural* view, for one would come to entertain it only after one's preconceptions had been sophisticated by knowledge of the set-theoretic antinomies, including not just Russell's paradox, but those of Cantor and Burali-Forti as well. (Boolos [13], p. 7, emphasis mine.)

The explanatory problem which underlies the violation of NATURALITY that Boolos touches on is a kind of vicious explanatory regress: the limitation of size conception explains, for instance, the universal set paradox by introducing a notion of oversizedness to set theory—which supports the explanation's DEPTH—and then showing that the universal set exceeds the established maximum possible set size. This in turn, however, demands an explanation of *why* the threshold condition is set at its level, and how this derives from the limitation of size conception. If the answer to this challenge is that the threshold is what it is on pain of contradiction, then the oversizedness explanation reduces to a syntactic argument, and so loses DEPTH. An independent explanation must then be found for the manifestation of the contradiction. But this reduces the notion of oversizedness to an idle wheel, and its explanatory purpose is left up in the air. We then have no explanation of the role of size in the notion of oversizedness for sets.

Soysal raises a similar point against the limitation of size conception and refers to it as an objection from the "explanatory insignificance of size."[8] Incurvati is likewise skeptical of the merits of limitation of size, but argues instead with Boolos that the conception cannot provide a "complete explanation" as it "depends on familiarity with substantial and sophisticated set-theoretic machinery."[9] In terms of the conditions introduced above, we can classify these complaints as arguments that limitation of size conception lacks respectively DEPTH and NATURALITY in its explanation of the paradoxes. However, Soysal's objection can also be seen to originate from the same violation of NATURALITY that Incurvati mentions. Namely, that without answering *why* size plays a natural role in set-hood, the resultant explanatory regress leads to the loss of DEPTH. Moreover, this issue cannot be

[7] This criterion is not explanatory but rather a precondition for the other virtues. See Subsection 1.2.3.

[8] Soysal [100], pp. 581-2.

[9] Incurvati [54], pp. 148-51.

addressed by merely pointing out the correlation of 'wide' sets like V and Ord with inconsistency. For example, Menzel [80] presents a natural variation of the iterative conception that accommodates atoms and for which the principle "for any cardinal number κ, there are at least κ atoms" holds.[10] Menzel's iterative conception thus contains a 'wide' set of urelements whose size exceeds any cardinality, even though V and Ord are still not sets on this account.

The application of NATURALITY sketched here also offers an answer to Paseau's comment, in reference to Boolos's demand for natural conceptions, that "Regarding ease of acquisition, it is hard to see why it should be thought important [to the correctness of a conception]."[11] We can now reply: from the above quote we can infer that part of what it means for Boolos to "readily acquire [a] conception" is that prior knowledge of the set-theoretic antinomies should not be necessary. The problem with the need for such conceptual preliminaries is that they may cause explanatory regress in the account of the paradoxes provided by the conception.[12]

3.2 Linnebo's Potentialist Account

Unlike the limitation of the size conception, all of the explanatory virtues introduced in the previous section have been argued to apply to the iterative conception. For NATURALITY we have already quoted Boolos. Similarly, Incurvati asserts that limitation of size contrasts with

> [...] the iterative conception, whose explanation of why there is no set of all ordinals or no universal set does not rest on the fact that the supposition that there are such sets leads to paradox (in the presence of Separation). (Incurvati [54], p. 150)

Furthermore, Incurvati [54] provides a general diagnosis of, and remedy to, the naive paradoxes from the iterative standpoint. This effort corresponds to our notions of DEPTH and UNIFORMITY. And with respect to MATHEMATICAL RICHNESS, Potter [89] offers a book-length argument for the justification of his set theory **SP** (Scott-Potter) on the iterative conception, which suffices for (almost) all mathematics.[13]

In the next subsection we discuss some issues that arise in formulating what kind of relationship exists between sets in the iterative conception. There appears

[10] This principles is inspired by Lewis's modal realism. See Menzel [80], pp. 57-8.
[11] Paseau [86], p. 40.
[12] I do not claim this is/would be Boolos's reply to Paseau.
[13] We discussed some caveats of this last claim in Section 1.1.

to be a sense of *priority* between some elements x, y, z and their set $\{x, y, z\}$, and clarifying this notion is essential to achieving DEPTH.[14] Traditionally, this priority has been understood either in the sense that some sets are constructed before others, or as a kind of metaphysical dependence.[15] We discuss arguments against both views in the next subsection. Then, in Subsections 3.2.2 and 3.2.3, we present Linnebo's alternative potentialist interpretation of set priority. Finally, we briefly discuss some problems of this approach in Subsection 3.2.4.

In Chapter 4, we come across the alternative which rejects that there is any sort of explainable priority relation between sets. This is the view of Boolos [13] and Incurvati [53] and naturally leads to minimalism about the iterative conception, which "holds that the content of the conception is exhausted by saying that sets are the objects that occur at one level or another of the cumulative hierarchy."[16] However, there I will argue that minimalism, unlike Linnebo's account, fails to provide a stable answer to why V is not a set (Section 4.5).

3.2.1 The Generative Metaphor and Ontological Dependence

The following is a common metaphoric description of the iterative conception, according to which sets are formed at different stages:[17]

> Indeed, every set is formed at some stage of the following "process": at stage 0 all possible collections of individuals are formed. [...] The sets formed at stage 1 are all possible collections of sets formed at stage 0 [...] The sets formed at stage 2 are all possible collections of sets formed at stages 0 and 1. [...] In general, for any natural number n, the sets formed at stage n are all possible collections of sets formed at stages earlier than n. [...] Immediately after all stages $0, 1, 2, \ldots$ there is a stage, stage ω. The sets formed at stage ω are, similarly, all possible collections of collections of sets formed at stages earlier than ω, i.e. stages $0, 1, 2, \ldots$. After stage ω comes stage $\omega + 1$: at which .... In general, for each α, the sets formed at stage α are all possible collections of sets formed at stages earlier than α. There is no last stage [...] (Boolos [13], p. 5)

[14] See e.g. Potter [89], p. 36.

[15] For example, Incurvati ([53], p. 73) writes: "We can distinguish two answers to this question: a *constructivist* one and a *platonist* one [rooted in metaphysical dependence]." (emphasis as in the original.)

[16] Incurvati [53], p. 84.

[17] The iterative conception was introduced in Subsection 0.2.

The sets in this *generative metaphor* are formed in phases of increasing combinatorical complexity, but exactly how this process should be interpreted is highly contentious. Since the non-structural axioms of **ZFC**, Pair, Union, Replacement, Separation, Choice, and Powerset, can be understood to act on available sets to form new sets, one obvious interpretation is to take the metaphor as a literal description of the iterative conception. The language used in the metaphor then suggests that the sets are formed in an actual "quasi-combinatorial" temporal process of creation, presumably carried out in thought by some faculty of the mind.[18] This provides an obvious interpretation of the supposed priority relation between sets, namely that x, y, z are prior to $\{x, y, z\}$ because those objects need to be available before the latter can be constructed from them, by applying Pair for instance.

A strictly constructive explanation is that the sets are literally formed step-by-step in time by some agent. Using some faculty of the mind I take x, y, z and create $\{x, y, z\}$ in my thought at some time t. This does demystify the generative aspect of set formation, but also intolerably restricts what sets we can have, as it is unclear even how countably infinite sets could be so created. If we allow some kind of *sub specie aeternitatis* perspective, where we can reflect on the process of generation itself, more sets may be reached, depending on the idealizations one accepts as constructivistic. Regardless, the assumed quasi-combinatorical powers need to be very strong to get anywhere. Already for $V_{\omega+1}$ you need to be able to generate sets that are incomputable or cannot even be defined in any language, since this level contains uncountably-many sets. Moreover, in $V_{\omega+1}$ there are, for example, set-theoretic ordinal representatives $(W_\alpha, \prec) \simeq (\alpha, <)$ for every $\alpha < \omega_1$, but there is no order that generates these sets, even as a supertask, because their cofinality is still ω_1, and no uncountable well-order embeds into $\mathbb{R}$ (interpreted as the time axis). Furthermore, any $\aleph_1$-subset of these representatives is isomorphic to ω_1, and therefore they too can be taken as representatives for ω_1. Hence there must either be ordinal representatives for ω_1 in $V_{\omega+1}$ that are generated simultaneously, which seems combinatorically unlikely, or their creation cannot be well-ordered in time.[19] The interpretation of the generative metaphor as an actual constructive process thus derails very quickly.

The alternative is to view the metaphor as just a convenient description which follows the standard narrative convention that (combinatorically) simple objects should be mentioned before more complex ones, if there is such a salient order.[20]

[18]The "quasiness" of the combinatorical picture of set formation refers to the generalization of the process for finite sets to infinite sets. The term originates with Bernays [9].

[19]This last argument against priority as a temporal modality is mine.

[20]Potter proposes another option:

We do not 'create' sets by applying axioms (notice the scare quotes), rather axioms are static guarantees that certain sets exist. Boolos [13] likewise makes the point that: "In any case, for the purpose of explaining the conception, the [generative] metaphor is thoroughly unnecessary."[21] But we can also understand set priority as a *dependence* phenomenon that does not arise between different constructions in time, but as an abstract ontological relation. The standard example here is that the existence of the set {Socrates} seems to depend on whether Socrates exists, while Socrates does not similarly depend on this singleton. The challenge is then to characterize this ontological dependence relation, since the obvious interpretation of x depends on y iff the existence of x necessarily implies that y exists, does not work. If Socrates exists, then {Socrates} necessarily also exists, because mathematical objects exist necessarily, if at all, according to standard accounts of metaphysical necessity. Incurvati ([53], pp. 75-9) discusses several analyses of the involved modality but concludes that "all the accounts under consideration seem to rely, at bottom, on some primitive notion of metaphysical dependence."[22] He then argues that our grasp of this primitive modality is shaky at best, and that there are good reasons to believe that it does not conform to the priority relation (even if there even is such a thing). It is not at all clear, for instance, that an entity cannot depend on itself, and Cameron [21] contends that the standard arguments against infinite descending chains of metaphysical dependence are unconvincing.[23]

3.2.2 The Potentialistic Conception

Apart from mental constructions and dependency, there is another angle to approach the generative metaphor. Linnebo [69] [68], Studd [102], Parsons [85] have emphasized the potentiality of the iterative hierarchy, not as a literal process or static dependence, but as open-endedness. Whatever sets there are, there can always be more:

> For given any attempt to pinpoint the extent of the hierarchy in a more definite way, it seems possible for the hierarchy to extend even further

---

> The difficulty the platonist now faces is to say what the relation of priority amounts to. One possible strategy which seems to be implicit in the thinking of many mathematicians is to regard platonism as a sort of limiting case of constructivism. (Potter [89], p. 38)

I won't comment further on this alternative strategy here.

[21] Boolos [13], p. 8. We used Boolos's metaphor-free description of the iterative conception in Subsection 2.2.1.

[22] Incurvati [53], p. 77

[23] Ferrier [32] gives two further arguments against priority as a form of dependence.

> [...] And since the hierarchy is supposed to extend as far as possible, this means that the proposed characterization cannot have been correct after all. So there seems to be something inherently potential about the set theoretic hierarchy. (Linnebo [68], p. 205)

This suggests a modal characterization of set formation. However, the involved modality cannot be the standard metaphysical one, because mathematical objects are typically deemed to exist necessarily, their existence being independent of spatiotemporal or causal concerns. Linnebo contends:

> This is not metaphysical modality in the usual post-Kripkean sense. Rather, the modality employed [...] is related to that involved in the ancient distinction between a potential and an actual infinity. This modality is tied to a process of building up larger and larger domains of mathematical objects. A claim is possible, in this sense, if it can be made to hold by a permissible extension of the mathematical ontology; and it is necessary if it holds under any permissible such extension. (Linnebo [68], pp. 207-8)

This "ancient distinction" concerns the nature of indefinitely extensible 'processes', and in particular in which cases these are completable or actualizable. The theme is already present in Euclid's Elements, as exemplified by the famous fifth postulate, which states:

> That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, *if produced indefinitely*, meet on that side on which are the angles less than the two right angles. (Euclid in Heath [45], Volume I, p. 155, emphasis mine.)

Here Euclid does not assert the static existence of certain lines within a figure under specific conditions. Rather, he posits the existence of a dynamic process, namely line extension with a straightedge, which, when executed by an ideal geometer, is assured to eventually culminate in an intersection.

Another example of such a process occurs with predicativism. Poincaré diagnosed the source of the set-theoretic paradoxes in the illegitimacy of referring to that which is being defined in its own definition. This suggests a potentiality which is similar to what is claimed of the iterative hierarchy: some things are defined first, then *these* are used to define more sets, and then *those* are used to define even more sets, etc., and this process can never be completed. We might then characterize predicative generation as follows: at the outset we recognize the

definiteness of the natural numbers, and so allow first-order quantification over them.[24] With respect to analysis this justifies all the real numbers (coded as subsets of $\mathbb{N}$) that can be defined arithmetically. Subsequently, we may proceed by admitting those reals that can be defined through quantification over the already accepted reals, etc. More precisely, we define the operation $S \mapsto S^*$ that takes the reals S and maps them to the collection S^* for which $X \in S^*$ iff

$$n \in X \leftrightarrow (\varphi(n))^S ,$$

where φ^S is any formula of second-order arithmetic with the second-order variables relativized to S, without free variables besides n. The predicative hierarchy then extends indefinitely under $(\cdot)^*$ and taking unions.

From these examples I isolate three general aspects any potentialist conception of a structure should specify:

1. The structures that the conception bootstraps from, the base stages, or initial worlds.

2. The generative operations which specify the objects that can be formed from those available. That is, the "permissible extension(s) of the mathematical ontology."[25]

3. The idealization and recombination conditions which lay down how and when the basic generative operations can be combined. These define the limit stages and the extent of the potential hierarchy.

These points illustrate that potentialism is not an absolute notion, but a framework which is capable of interpreting and unfolding different mathematical structures against various foundational stances. Feferman argues for a similar perspective concerning predicativity:

> So if one accepts this more general standpoint [i.e. the legitimacy of predicative unfolding against various backgrounds], one answer to the question "What is predicativity?" is that it is a concept applicable to different foundational stances given by the rejection of the actual infinite for various domains, coupled with its possible limited acceptance

[24] As a justification we could point out, for example, that some infinitary reasoning in analysis is indispensable to science. Alternatively, we may agree with Poincaré and Weyl that "the natural number system and the associated principle of induction constitute an irreducible minimum of theoretical mathematics, and any effort to 'justify' that would implicitly involve its assumption elsewhere." (Feferman, [29], p. 9).

[25] Linnebo [68], p. 208

> for others. The logical problem in each case is to characterize exactly the limits of that particular stance. The potential value for philosophy then is to be able to say in sharper terms what arguments may be mounted for or against taking such a stance. (Feferman [29], p. 29)

3.2.3 Modal Set Theory

Linnebo motivates his modal plural set theory by interpreting the indefinite extensibility of the sets through a Kripke model (W, R, D), where one world, or possible stage of the process of set formation, $w' \in W$ is regarded as accessible from another $w \in W$ (wRw'), just in case one can get from w to w' by "a legitimate expansion of the mathematical ontology" from $D(w)$ to $D(w')$.[26] Thus wRw' implies $D(w) \subset D(w')$, and so we will write $w \leq w'$ instead of wRw'.

This partial order can be further characterized: set formation on the iterative conception is well-founded. Furthermore, if some sets and some other sets possibly exist, then all of them also possibly exist. Hence $\leq$ is a directed partial order, and so the modal logic **S4.2** is sound with respect to the intended system of possible worlds. The modality $\Diamond$ here signifies possibility in a legitimate expansion of the ontology. Thus $\Diamond\varphi$ holds iff there is such an evolution where φ is true. Linnebo further argues that membership of a plurality, denoted $\prec$, is *rigid* and *inextensible* across possible worlds.[27] He expresses rigidity as

$$u \prec xx \to \Box(u \prec xx)$$
$$u \not\prec xx \to \Box(u \not\prec xx)\,,$$

where u is a set and xx a plurality; and the idea that plurals are inextensible is captured by relativizing the Barcan formula $\forall u \Box\theta \to \Box\forall u\theta$, which corresponds to the semantic requirement that objects do not disappear between worlds (domains do not shirk), to the condition $u \prec xx$. That is,

$$\forall u(u \prec xx \to \Box\theta) \to \Box\forall u(u \prec xx \to \theta)\,,$$

for any formula θ. This motivates the adoption of following background modal plural first-order logic:

Definition 2 (MPFO). The system **PFO** extends any system for first-order logic with the usual introduction and elimination rules for the plural quantifiers and the plural comprehension axiom

$$\exists xx \forall u(u \prec xx \leftrightarrow \varphi(u))\,.$$

[26] Linnebo [68], p. 208

[27] In [71] Linnebo discusses several arguments for the rigidity of plurals.

The system **MPFO** turns **PFO** into a modal logic by adding the axioms of **S4.2**, the rigidity axioms, and all instances of the inextensibility schema.

Leveraging the inextensibility of plurals we can capture the *extensional definiteness* of a formula φ in **MPFO** as

$$\exists xx \Box \forall u (u \prec xx \leftrightarrow \varphi(u)) .$$

Linnebo's most significant contribution in [69] is his Interpretation Theorem, which bridges deductibility in the ordinary and modal setting. To state this result, we need to define a translation from the non-modal language to the modal one:

Definition 3 (The Potentialist Translation)**.** If a quantifier $\forall, \exists$ occurs in a formula as $\Box\forall$ or $\Diamond\exists$, we say that the quantifier is *modalized*. When all quantifiers in a formula are modalized, we refer to the formula as *fully modalized*. Given a non-modal formula φ, its *potentialist translation* $\varphi^\Diamond$ is the fully modalized formula obtained by replacing every ordinary quantifier in φ with the corresponding modalized quantifier.

Theorem 1 (Linnebo's Interpretation Theorem)**.** *If $\vdash$ is the classical deductibility relation (without plural comprehension) and $\vdash^\Diamond$ comprises $\vdash$,* **S4.2***, and the rigidity axioms for every atomic predication in the language, then for any non-modal formulas $\theta_1, \ldots, \theta_n$ we have*

$$\theta_1, \ldots, \theta_n \vdash \varphi \quad \text{iff} \quad \theta_1^\Diamond, \ldots, \theta_n^\Diamond \vdash^\Diamond \varphi^\Diamond .$$

Proof. By induction on proofs. □

The significance of this theorem is that it

> [underlies] the observation [...] that modal set theory is compatible with ordinary nonmodal **ZF** set theory but looks at the same subject matter under a 'finer resolution' [...] that is made available [...] [by formulas which are not fully modalized,] to enable arguments and explanations that are not available in ordinary nonmodal set theory. (Linnebo [68], p. 214)

Moreover, the theorem confirms that the modalized quantifiers behave logically just like ordinary quantifiers. This lends credibility to Linnebo's idea that we have implicitly operated with modalized quantifiers all along, albeit without recognizing the potential for a finer resolution when decoupling the modalities.

The main application of this modal decoupling that is discussed in [68] is the reinterpretation of Cantor's naive set existence principle in the modal setting. The principle can be rephrased as:

$$\Box\forall xx\Diamond\exists y\Box\forall u(u \in y \leftrightarrow u \prec xx)\,.$$

This says that for every plurality xx it is possible to form a set y so that no matter the evolution of the open-ended set formation, y and xx will be co-extensional. It can then be shown that this principle implies exactly those fully modalized instances of Unrestricted Comprehension for which the comprehension formula φ is possibly extensionally definite, or

$$\Diamond\exists xx\Box\forall u(u \prec xx \leftrightarrow \varphi(u))\,.$$

In other words, those instances of $\mathrm{UC}^{\Diamond}$ are obtained for which φ is not *indefinitely extensible*, in the sense that

$$\forall xx\Diamond\exists u(u \not\prec xx \leftrightarrow \varphi(u))\,.$$

This Cantorian analysis blocks the set-theoretic paradoxes, however "the key notion [here] of a condition's possibly being extensionally definite is not the potentialist translation of any non-modal formula and is thus expressible only in the richer modal framework."[28]

Using the established framework, Linnebo argues that set membership $\in$ is extensionally definite, and that the well-foundedness of the worlds under $\leq$ justifies Foundation in the individual worlds.[29] Furthermore, he contends that the principle of Maximality—according to which all the sets that may be formed are formed as soon as possible—implies the extensional definiteness of the subset relation $\subset$.[30] This suggests the following basal modal set theory:

Definition 4. Let $\mathbf{L}_{\text{base}}$ be the theory that adds to **MPFO** the modal Cantor principle, the extensional definiteness of $\in$ and $\subset$, Extensionality, and Foundation.

We then have the following recovery theorem:

Theorem 2.

(a) $\mathbf{L}_{\text{base}}$ *proves the potentialist translations of all the axioms of* $\mathbf{Z}-$ Infinity.

(b) $\mathbf{Z}-$ Infinity *is interpretable in* $\mathbf{L}_{\text{base}}$.

[28] Linnebo [68], p. 220.

[29] See Linnebo [68], section 6.1 and 6.2.

[30] Op. cit. section 6.4.

Using two further hypothesis about the nature of the set theoretic modality we can extend this result to **ZF**. One reasonably natural assumption is that if φ is extensionally definite and every φ is correlated with a unique ψ, then ψ too is extensionally definite. In other words,

Definition 5 (ED-Repl)**.** Let $\mathrm{F}(\psi^{\Diamond}(u,v))$ abbreviate that $\psi^{\Diamond}(u,v)$ is a functional formula, in the sense that

$$\Box\forall u\Diamond\exists v\Box\forall v'(\psi^{\Diamond}(u,v') \leftrightarrow v = v')$$

Then ED-Repl is the axiom

$$\mathrm{F}(\psi^{\Diamond}(u,v)) \rightarrow \Box\forall xx\Diamond\exists yy(\forall u \prec xx)(\exists v \prec yy)\psi^{\Diamond}(u,v)\,,$$

which says that if $\psi^{\Diamond}$ is functional, then for any plurality xx that may arise, there will be the plurality yy of those sets in the image $\psi^{\Diamond}(xx)$.

The other plausible assumption we may entertain is that truths about the potential hierarchy are mirrored in truths about individual possible worlds. This suggests the following modal reflection principle:

$$\varphi^{\Diamond} \rightarrow \Diamond\varphi\,. \qquad (\Diamond\text{-Refl})$$

The intuition here is that if the quantifiers of a formula ψ range over all possible sets, so that it is of the form $\varphi^{\Diamond}$, and ψ holds, then φ must be genuinely possible, in the sense that $\Diamond\varphi$.[31] This reflection principle is another instance of the finer resolution that the modal logic provides.

These two hypothesis correspond to the following theorems:

Theorem 3. $\mathbf{L}_{\text{base}}$ + ED-Repl *entails the potentialist translation of Replacement.*

Theorem 4. $\mathbf{L}_{\text{base}} + \Diamond$-Refl *entails the potentialist translation of Infinity.*

Let **L** be the theory $\mathbf{L}_{\text{base}}$ + ED-Repl + $\Diamond$-Refl, then

Theorem 5. *Non-modal* **ZF** *is interpretable in* **L***.*

Finally, we have the converse:

Theorem 6. *The modal set theory* **L** *is interpretable in* **ZF***. Hence* **L** *and* **ZF** *are equiconsistent.*

[31] For technical reasons we should not allow φ to contain plural quantifiers. See Linnebo [68], pp. 222-3.

3.2.4 Some Problems with Linnebo's Potentialism

We briefly state some challenges that Linnebo's modal set theory **L** faces. Unfortunately, we cannot delve further into these problems as they are beyond the scope of this thesis.

Convergent Potentialism is Implicit Actualism

Hamkins [50] contends that the "convergent potentialism" which is implied by the underlying modal logic **S4.2** is "implicitly actualist." He argues that genuine possibility in this context ought to be interpreted more radically as the potential to irreconcilably branch the universe. This consideration favors a theory founded on **S4**, which breaks Linnebo's Interpretation Theorem.[32] Hamkins writes:

> [...] I regard convergent potentialism as implicitly actualist, [...] if the universe fragments of one's potentialist system are mutually coherent with one another [...] then there is a unique limit model to which the system is converging, and there seems to be very little at stake in the ontological dispute concerning the actual existence of this limit model. What does it matter if the potential objects that might come to exist do not yet actually exist, if the way that they will come to exist is unique and deterministic? To experience potentiality in convergent potentialism is simply to wait for the inevitable. In a convergent potentialist system, after all, the potentialist can refer to truth in the limit model, without having that model in his or her ontology, by means of the potentialist translation. (Hamkins [50], p. 33)

This echoes the Cantorian sentiment that potential infinities presuppose actual infinities. Brauer [20] takes up Hamkins's idea but argues that **S4** is not sufficiently expressive and needs to be complemented by an "inevitability operator." He then shows that the resultant modal logic **S4**+ is, however, not axiomatizable. This outcome sheds significant doubt on whether "there is any natural translation of ordinary mathematical discourse into the potentialist framework in the context of branching possibilities."

The Merit of Modal Reflection

The base system $\mathbf{L}_{\text{base}}$ of Definition 4 is unable to derive Infinity$^{\Diamond}$. Therefore Linnebo introduces $\Diamond$-Refl as "another natural principle" about the potential hi-

[32] See Hamkins [50], pp. 32-38, the introduction section of Hamkins & Linnebo [52], and the paper Linnebo & Shapiro [72] about different forms of potentialism, especially section 7.

erarchy to derive it from. This principle mimics the **ZF** reflection theorem

$$\varphi \to \exists\alpha\varphi^{V_\alpha}\,.$$

Furthermore, the proof that $\Diamond$-Refl implies Infinity$^\Diamond$ is analogous to the proof that the reflection theorem implies Infinity.[33]

Technically, $\Diamond$-Refl does what it needs to do, but so would just adding in Infinity$^\Diamond$ as a further assumption. The question is, however, whether $\Diamond$-Refl has any merit over accepting modalized Infinity directly, for instance that it is easier to justify on the potentialist conception or is explanatorily superior. I doubt that such an argument will be easy to make. The modal reflection principle does not yield any new theorems, and certainly in the non-modal setting we accept Infinity over the Lévy–Montague Reflection Theorem as a more basic truth. But if we can just as well drop in Infinity$^\Diamond$, we lose an important exemplification of the finer resolution approach, and so we may then wonder whether we are doing more than only pushing around boxes and diamonds.

The Nature of the Set-Theoretic Modality

Although the intuition for **L** was developed above using Kripkean possible world semantics, the interpretation of the set theoretic modality behind the open-endedness of sets cannot be in terms of metaphysical possibility. Linnebo contends:

> Parsons and I seek to model a sense in which the existence of each individual set is potential relative to the existence of its elements and consequently the hierarchy of sets itself is inherently potential. Although further work is no doubt needed to clarify the kind of modality involved in these ideas, both authors are very clear that the relevant modality is not the metaphysical one. (Linnebo [68], p. 226)

However, a clarification of this kind remains lacking.

Putnam [93] takes the modality to be a kind of mathematical possibility, in the sense that we may specify a set-theoretic universe, but that this specification is always subject to mathematical extension. This modality should be understood in terms of Putnam's *objects-modalities duality*, which he compares to the wave-particle duality in quantum mechanics. It provides an alternative lens through which to perceive the familiar object-situation, because, as Putnam emphasizes, there is no significant *mathematical* difference between the assertion that such and

[33] See Theorem 8.5 in Linnebo [68] and, e.g., Theorem II.5.3 in Kunen [57], p. 131.

so objects exist and the assertion that it is possible to select objects that are such and so.

Linnebo contends that "[Putnam] points in the right direction, namely that mathematical objects are a kind of reified possibilities."[34] However, a challenge for Putnam's view that Linnebo raises is that the non-mathematical circumstances may be perversely intertwined with the mathematical context. Suppose, for example, that it is possible for there to be a certain set, but only in those worlds where there is enough stuff of some type. Then this set cannot be treated within a purely mathematical modal set theory, and we are prevented from freely working with it. The response that Linnebo proposes is to decouple the modality into a *circumstantial* modality, which keeps the interpretation of the language fixed but allows the non-mathematical circumstances to vary, and an *interpretational* modality, which works in the reverse. Putnam's mathematical modality can then be understood as the interpretational axis in this multi-modal setting. Evidently, this analysis still requires substantial work. In particular, it needs to clarify the nature of the "reified possibilities" and the sense in which the language is supposed to vary.[35]

3.3 Conclusion

I have argued that Boolos's naturality condition can be understood as the demand that our conception of set should be able to answer certain why-questions. These considerations favor the iterative conception over, for instance, the limitation of size conception, because the latter lacks the NATURALITY virtue in its explanation of the paradoxes (Section 3.1). To attain DEPTH, we also need to elucidate the sense of priority that exists between some elements x, y, z and their set $\{x, y, z\}$. We saw that neither the generative metaphor nor the metaphysical dependency interpretation offered a satisfactory clarification of this notion (Subsection 3.2.1). We then introduced in Subsection 3.2.3 Linnebo's potentialist account of the iterative conception and his modal plural set theory. This account possesses NATURALITY and UNIFORMITY, and provides a plausible interpretation of the priority relation. Theorem 5 furthermore establishes MATHEMATICAL RICHNESS. Regarding DEPTH, Linnebo contends:

> Of course, one may wonder whether potentialists too are under pressure to concede that the set theoretic hierarchy corresponds to a set, which would show their view to be unstable. But potentialists have a

[34] Linnebo [70], p. 264.

[35] Ferrier ([32], section 4) presents several other arguments against the modal view.

> response. On their conception, the hierarchy is potential in character and thus intrinsically different from sets, each of which is completed and thus actual rather than potential. This intrinsic difference affords potentialists—unlike their absolutists rivals—a reason to disallow the disputed set formation. (Linnebo [68], p. 206)

In the next chapter we will look at the status of SIMPLICITY. In particular we will discuss whether potentialism has any explanatory advantages over a more minimalist approach offered by Soysal [100]. My conclusion will be that this is indeed the case, and that therefore potentialism provides a good explanation of the paradoxes based on the criteria we have set out.

Chapter 4
Potentialism versus Minimalism

4.1 Idle Wheel or Resolution Magnifier

We have seen in the previous chapter that potentialism provides a plausible and uniform account of the set-theoretic paradoxes which complies with the criteria set forth in Section 3.1. However, Soysal [100] has argued in recent work that the involved modal apparatus is merely an "idle wheel":

> If, moreover, we are given no independent grip on potentialist's notion of modality (because we are told it is primitive and idiosyncratic to set theory), then what stops us from simply interpreting the domains of the worlds w_α [of the Kripke model reflecting the modality] as stages V_α defined in **ZFC**? What exactly is added by the '$\Box$' and '$\Diamond$' in front of quantifiers? Potentialism on this option starts to look like a notational variant of set theory. And this surely affects its explanatory power. (Soysal [100], p. 588)

Soysal's point is that the modal machinery on which Linnebo relies can only be afforded if, for example, it provides an explanatorily superior account of the paradoxes or justifies a stronger theory of set. Linnebo conceives of his potentialism not as an alternative to the iterative conception and its **ZFC**-centered set-theoretic practice, but rather as an elucidation of the underlying mechanics. The increased *resolution* that is so obtained is codified in a modal language and then deployed to confront foundational questions:

> The picture that emerges [...] is that modal set theory is a valuable supplement to standard nonmodal **ZF** set theory rather than a rival. The modal theories provide powerful instruments for studying the same subject matter under a finer resolution. For the purposes of almost all of ordinary mathematics and much of set theory, this finer resolution is not needed. In these cases no insight is lost, but great expressive and inferential ease is gained, by relying on standard nonmodal set theory. *However, when we confront difficult foundational and conceptual questions concerning set theory, the finer resolution provided by the modal approach can be very valuable.* (Linnebo [68], p. 206, emphasis mine.)

Linnebo's modal potentialist approach to the iterative conception thus stands and falls with its ability to contribute to "difficult foundational and conceptual questions," such as those posed in Problem 1. We focus in this chapter on the explanatory power of potentialism with regard to the set-theoretic paradoxes. Following Soysal, we single out the universal set paradox and the associated *why-question*: why is the totality of all sets in the iterative hierarchy not itself again a set?[1] The upcoming discussion will however not specifically depend on this example. The other paradoxes can be treated analogously.

To challenge the explanatory merit of the potentialist position, Soysal proposes a rival answer to the why-question. For Soysal V is not a set because the iterative conception, for instance, sanctions Foundation and Pair, and selfcontaining sets like V are directly at odds with those principles. That is, the universal set paradox is just the result of incompatible conceptual requirements, *and there is no more to it than that.* Soysal classifies this "conception-based" explanation as *minimalist* in the sense that it does not rely on further (metaphysical) premises beyond whatever is necessary to make sense of the iterative conception.[2] She then argues that explanations which are not minimal like this, such as those based on limitation of size and potentialism, still "rely on the minimal explanation and [yet] add nothing to it."[3]

If we would find Soysal's minimalist account persuasive, then this would take the wind out of Linnebo's sails. Her conception-based explanation is maximally parsimonious, so if it went through that would make a good case that the potentialist's modalities are indeed mere idle wheels, at least in so far as explanatory power is concerned.

In Section 4.2 we discuss the critique voiced by Yablo [109], Studd [102], and Dummett ([28], chapter 24) that minimalism does nothing more than simply point out contradiction and therefore is uninformative. We also consider some obvious but unsatisfactory minimalist responses. In Section 4.3 we turn to Soysal's conception-based rebuttal of this criticism. I will argue against Soysal that her response fails to make proper distinction between the intensional and extensional viewpoint, and so misses the main point of the critique against minimalism. We will see that this failure ties into the question of *what* V is. Is it a set or set-like collection that is just not part of the cumulative hierarchy, or is V not set-like at all?

---

[1] Soysal [100], pp. 575-576.

[2] This includes the idea that iterative sets are somehow dependent on their members. Boolos ([13], p. 8), for instance, argues that: "In any case, for the purpose of explaining the [iterative] conception, the metaphor [of sets being prior to their members] is thoroughly unnecessary [...]"

[3] Soysal [100], p. 591.

If we take V as a non-set, set-like collection we run into the classical problem of the set/class distinction. But as Boolos remarked: "We ought to be able to formulate a theory that reflects our beliefs [about sets as general objects of collection]."[4] As to viewing V as a non-iterative set, I will present what I call a *conception extension argument* to show that this leads to *conceptual instability*.

Regarding the other option, namely taking V as not set-like at all, I propose a reformulation of the conception-based minimalist answer to the why-question: V is not a set because it is not a set on the iterative conception *and the iterative conception is correct*. It seems that Soysal also has this approach in mind, but I will argue that she fails to recognize the impact which this implicit clause has on the minimalist position: it tacitly loads the weight of the explanation onto the correctness assumption. The problem is that whatever argument we give for the correctness of the iterative conception, it becomes implicitly incorporated into this explanation. So for this to not undermine the minimalist, the correctness argument cannot depend on any assumption that is not part of whatever is minimally required to grasp the iterative conception.

In Section 4.4 we turn to whether this hard bargain can be met by Incurvati's minimalist account of correctness, which proceeds by *inference to the best conception*. However, in Section 4.5, I will argue that even if this approach is successful, at best it only picks out a *family* of conceptions of set, which all share the virtuous aspects of the cumulative hierarchy. I then show that this leads to *explanatory instability* because there are conceptions in this family that provide diametrically opposed answers to Soysal's why-question. Finally, I conclude that in contrast to the conception-based minimalist explanation, potentialism does track a specific feature of the world, namely potentiality, which not only explains *why* V is not a set, but also clarifies *what* it is instead.

4.2 Critique of Minimalism

The paradoxes we discussed in Subsection 1.2.1 show that we can run into contradiction even when we try to talk about collections such as V and Ord, which naturally arise in our practice and for which we have some intuition about their presumed membership. When this happens we would like to have an explanation of what it is about their putative members that fundamentally prevents them from aggregating as a set, not just a demonstration that the supposed set has been banned by our theory. As Dummett ([28], p. 316) puts it regarding the Cantor paradox: "to [merely] say, 'If you persist in talking about the number of all cardi-

[4]Boolos [15], p. 441.

nal numbers, you will run into contradiction', is to wield the big stick, not to offer an explanation." What we want is to identify some feature that explains "what it is about the world that allows some sets to form a set, whilst prohibiting others from doing the same."[5]

One may wonder what more could possibly be demanded from an explanation of why a certain set cannot exist than a proof in our best theory that its presumed existence leads to contradiction. After all, in many other cases this appears to be sufficient. For instance, it seems to me that an acceptable explanation of why there cannot be a barber who cuts all people that do not cut themselves, is to simply say that such a barber would have to do contradictory things, namely cut himself and not cut himself. Nothing more needs to be said here.

There is, however, a crucial difference between sets and barbers: we do not have any pre-theoretical intuition about what kind of barbers should exist. Our understanding of a 'diagonalizing' barber derives purely from the description of such a barber, so no similar conflict of intuitions arises. On the other hand, we *do* have a pre-theoretical grasp of the concept of set, and it seems that

> [...] where the set's putative membership is clear, the [quasi] *de re* explanation is more revealing [than a *de dicto* explanation]. One wants to know not why a set thus and so specified cannot exist, but why a set of such and such a nature—with such and such members—cannot exist. (Yablo [109], p. 148, emphasis as in the original.)

Likewise, Studd writes:

> However, the question of real interest is not why this instance of Naïve [Unrestricted] Comprehension yields a contradiction, but why certain sets—in this case, those that lack themselves as elements—are unable to form a set. And this cannot be explained merely by appeal to logical truths. [...] The derivation of a contradiction from the relevant instance of [Unrestricted] Comprehension shows that [the sets z such that $z \notin z$] do not form a set, but fails to explain why they cannot. (Studd [102], p. 700)

The explanatory problem should thus be understood *conceptually*, not logically. It is not that we want to explain why an intensional—or in Yablo's words, *de dicto*—description of some pathological set contradicts our iterative set theory. Rather, we want to untangle, from an extensional viewpoint, what it is about the members of that putative set which conflicts with the intuition we have in our naive

[5] Studd [102], p. 699.

concept of set that, whenever the members $a, b, c, \ldots$ of a plurality are determinate and exist, those members should be able to form the set $\{a, b, c, \ldots\}$.[6] Therefore, although the Russell set R simply cannot be a set as a consequence of pure logic, much like Russell's barber cannot exist, this does not explain what is wrong with the intuitive principle that for every plurality there is a co-extensive set, which *does* guarantee the existence of R. Linnebo phrases this puzzle as follows:

> We are thus confronted with a paradox, in the sense that we have made a number of assumptions, all of which seem perfectly plausible, but which jointly lead to contradiction. [...] The semantics of plural logic ensures that a plurality consists of a determinate range of objects. But a set is completely characterized by its elements. A plurality thus provides the resources for a complete and precise characterization of a set. *So what could prevent us from collecting the given plurality into a set?* (Linnebo [69], p. 147-9, emphasis is mine.)

Let us call this the *Collapse Problem*.[7]

A minimalist might counter to Linnebo's question that our naive view of sets is simply wrong, and that we now accept Foundation as part of our theory. This uniformly blocks the pathological instances of Unrestricted Comprehension. But to block the paradoxes is one thing, and to explain them quite another. Foundation is not part of our (naive) concept of set, so it needs independent justification if it is to serve an explanatory purpose. The risk is otherwise that we might start to confuse cause and effect, as it is possible that part of the reason why we accept Foundation in the first place, is to keep pathological sets out. Yablo similarly contends that:

> [...] one may question whether [well-foundedness] is basic enough. To say that V cannot exist because it would be ill-founded seems to get things the wrong way around. It is because sets like V are independently problematic that we are drawn to a requirement that keeps those sets out. (Yablo, [109], p. 149)

That a certain factor X, such as ill-foundedness, *correlates* with the paradoxes, also does not by itself establish that X *causes* trouble. The pathological sets are

[6] Yablo [109], p. 148, for instance, writes that: "A set-theoretic paradox arises when a set that, naively speaking, ought to exist cannot exist, because the hypothesis of its existence conflicts with fundamental facts about sets. An explanation of the paradox is an explanation of how and why the conflict arises."

[7] Linnebo [69] refers to the involved principle as COLLAPSE. It is equivalent to Unrestricted Comprehension over standard plural logic.

also, for instance, all infinite, but we do not want to say that infinite sets are pathological because of this (well, *we* do not). Thus, regardless of whether or not we accept Foundation in part to keep paradox out, the point is that in naive set theory Unrestricted Comprehension (UC) leads to inconsistency. It is *this* that needs explanation: why is it that this perfectly natural view of sets yields absurdity? Switching to a different conception where the problematic instances of UC (or, COLLAPSE, for Linnebo) are unavailable, is not helpful if the principles that block them cannot be independently argued for.

Without proper analysis of the paradoxes *within the context in which they originally arose*, there is no assurance that a foreign principle such as Foundation, while it can be used to block the paradoxes, does any explanatory work that contributes to the coherence of the system. If we are not mindful of this problem we may turn out to be like the person who denies the need for vaccination simply because there are no current infections in the area. However, the underlying pathogen could still be widespread and potentially wreak havoc at any point in the future.

4.3 Conception-Based Minimalism

According to Soysal, however, the minimalist approach to explanation can do more to address the why-question than simply point out the theorem $\ulcorner \mathbf{ZFC} + V \text{ exists} \vdash \bot \urcorner$. She agrees that this observation lacks any specific insight into the sets because it "fails to provide the deeper reason why these axioms that prohibit the existence of the universal set are there in the first place."[8] Instead, Soysal proposes that, accepting the correctness of the iterative conception—and according to her the philosophers that ask the why-question do just that—the non-set-hood of V is explained by the structure of the cumulative hierarchy, since this structure *sanctions*, e.g., Foundation and Pair:

> The universe of sets is not a set because the supposition that it is contradicts some axioms of **ZFC** [e.g. Foundation and Pair], *and these axioms are part of the iterative conception of sets.* (Soysal [100], p. 592, emphasis is mine.)

So *that* is the deeper reason Yablo and Studd find wanting, and there are no "sticks" involved:

> In particular, what [this] conception-based explanation adds to [the] minimal explanation [criticized by Yablo and Studd] is the claim that

[8] Soysal [100], pp. 577.

> Foundation and Pairing are part of the iterative conception of sets. The conception-based explanation is thus a "deep" explanation [...] *It provides an independent reason for why the axioms that prohibit the existence of the universal set are there in the first place: They are simply part of our iterative conception of sets.* (Soysal [100], p. 592, emphasis is mine.)

There are, however, two immediate problems with this explanation. First, Soysal still answers the why-question with an intensional interpretation of V in mind, or at least she does not seem to appreciate the importance of distinguishing this view from the extensional viewpoint. On the contrary, I have argued that the Collapse Problem and the common critiques of minimalism, which we discussed in Section 4.2, adopt a perspective of the paradoxical sets as 'lassoed' pluralities. After all, sets are an extensional species and should therefore be treated as such whenever their putative membership is understood. This is precisely Yablo's point.

That Soysal's intensional interpretation does not conform with the intended extensional view is evident from its inability to differentiate between the logical and conceptual accounts of Russell's Paradox. The former boils down to the theorem $\ulcorner R$ exists $\vdash \bot \urcorner$, and the latter, on Soysal's take, to "the Russell set is not a set because the supposition that it is contradicts logic, and logic is part of the iterative conception of sets." It is clear that neither explanation—if there is any difference at all—is an acceptable answer to the Collapse Problem.

A second and more important issue is that such a response to the why-question gives us no information about the what-question: "What, if not a set, is the universe of sets?"[9] There are two options here:[10]

1. Accept that we can form a collection (some set-like object) of all iterative sets, but deny that this is itself an iterative set. That is, deny that it is in the cumulative hierarchy.

2. Deny that we can form a collection of all iterative sets *simpliciter*. That is, deny that there is a mathematical object which is set-like and has all and only the iterative sets as members.

On Option 1 the Soysalian minimalist could try to sidestep the Collapse Problem by saying that her minimalism is only supposed to be about the iterative sets.

[9] Soysal [100], p. 575.

[10] Alternatively, some philosophers have opted to reject pluralities altogether. We disregard this option for two main reasons: 1) Our dialectical stance necessitates the acceptance of pluralities; otherwise, we wouldn't be able to analyze Linnebo's program. 2) Soysal accepts pluralities.

She could then explain: "V is not a set on the iterative conception because there is no stage at which it could have been formed. This is a consequence of the well-orderedness of stages, which is part of the conception. With regard to the Collapse Problem, however, I am happy to grant that V is a non-iterative set or an extensional class of some sort (a set-like, non-set collection), but I am simply not concerned with such entities." This response seems adequate and yet minimal with respect to the iterative conception, but it does imply commitment either to classes or sets that are not in the cumulative hierarchy.

Commitment to classes is, however, controversial. The pre-theoretic belief is no doubt that there is only a single category of collection, namely just sets. The notion of classes as set-like objects which are distinct from the sets arises only after encountering the paradoxes, rather than being an inherent feature of the theory of collection itself. This violates the general desideratum of theory formation that theories ought to treat their subject matter with the kind of uniformity that is suggested by our pre-theoretic beliefs, or as Boolos puts it: "Set theory is supposed to be a theory about *all* set-like objects."[11]

There are also various direct objections to classes considered as independent objects. For example, based on his analysis of the use of classes in natural language, Boolos [15] [16] famously argued that while there are all the sets, classes are ontologically innocent plural locutions. Hence, talk about classes can always be dispensed with, and therefore does not lead to paradox. Parsons [84], on the other hand, links the systematic ambiguity he perceives in the language of set theory and the problem of quantification over *all* sets to the potentialist idea that "a perspective is always possible according to which your classes are really sets."[12] Both arguments, however, reject classes as freestanding entities which would justify a class-based answer to Option 1.

What about non-iterative sets? This is also problematic, especially for the minimalist. Suppose there is a set V of all iterative sets, then we could formulate a new conception of set, whose story might unfold as follows: There are the null set and the set containing just the null set, sets of all those, sets of all *those*, sets of all *Those*, ... There are also sets of all ***THOSE***. Let us now refer to these sets as "those." Then there are sets of those, sets of *those*, etc. Along with all these sets that we get from iterating 'set of' starting from the null set (the sets in the standard cumulative hierarchy), there is also the set V of all of them. Then there

[11]Boolos [18], p. 442, emphasis is mine. Also in [15], p. 441, Boolos asserts: "We ought to be able to formulate a theory that reflects our beliefs." Tait ([103], p. 10) even calls the supposed distinction between sets and classes "an embarrassment in set theory."

[12]Parsons [84], p. 11. See also Boolos's [17] direct reply to Parsons, and Linnebo's [67] discussion of their respective positions.

are sets of all *these*, etc., analogous to how we iterated from the null set.[13]

This conception of set appears to be just as iterative as the iterative conception itself, which covers the story up to "Along with. . . ." So, we find ourselves in a situation reminiscent of the problem of classes discussed earlier. Specifically, there are objects of a certain kind (iterative sets) that, under pressure of paradox, we are forced to sort into distinct categories (iterative sets starting from either $\emptyset$ or V). But, again, there is no such pre-theoretical intuition, and both conceptions ostensibly describe the same things. After all, the cumulative hierarchy is supposed to include *all* iterative sets. Thus we have a *conception extension argument* which shows that Option 1 leads to *conceptual instability*.[14]

The only way I see out of this is to deny that one can take arbitrary subsets of V, so that subsequent layers of sets are not of full rank and are therefore distinctly separate from the earlier stages. Two conceptions that implement suchlike approach are Forster's [34] multi-wand generalization of the iterative conception and Quine's system **NF**.[15] However, **NF** denies Separation in general and therefore does not support an iterative conception of set below V either. And it is hard to phantom how one could possibly explain the failure of Separation on V without drawing on some presumed property of the iterative process, which ensures the special status of V as a set comprised of iterative sets that is itself nevertheless non-iterative. But this is already out of reach for the minimalist, who cannot accept such a property as it is certainly not required to grasp the iterative conception (what goes on *in* V).

We now turn to Option 2, which Soysal also seems to have in mind. She argues that implicit in what we mean by set—our concept of set—when we ask the why-question is the iterative conception:

> In my view, the iterative conception of sets is part of the [. . .] concept

[13]Adapted from Boolos [13], p. 8.

[14]Studd ([102], p. 700) defends potentialism in a manner that is reminiscent of this style of argument: suppose that V corresponds to the members of some inaccessible rank V_{θ_0}. It seems that there *could* have been a bigger universe V_{θ_1}, where θ_1 is some inaccessible larger than θ_0. But then the actual universe $V = V_{\theta_0}$ could have been a set, which can't be. So, "to explain why these sets are unable to form a set, it would need to be supplemented by an explanation of why they—viz., the elements of V_{θ_0}—must be all sets there are. [. . .] Certainly *this* is no logical truth." The issue I have with this argument is that it seems to presuppose something which is already quite close to potentialism. To think that V can be represented as V_{θ_0}, one must accept a perspective of the universe as something that can be looked at 'from the outside'. An actualist does not have to agree that this view can be made sense of.

[15]We will return in the next section to Forster's conception of set. Incurvati ([54], chapter 6) argues that **NF** is not a conception of set, but rather a conception of objectified property. Boolos ([12], p. 219) disregards the theory as "unnatural."

> of set; it is part of what we—we who ask the why-question and work with the iterative conception—mean by 'set'. On this view, not only are Foundation, Pairing, and Separation part of the iterative conception of sets, they are also part of the concept of set. (Soysal [100], pp. 592,5)

This suggests the following reformulation of Soysal's conception-based explanation: V is not a set because it is not a set on the iterative conception (due to e.g. Foundation and Pair), and the iterative conception is correct (there are no non-iterative sets). The Collapse Problem is now also easily answered because the sets just do not arise as singularizations of pluralities, but by iteration. The Russell set and V are simply not in the cumulative hierarchy, and the sets in this hierarchy are all the sets there are.[16]

Formulated like this, it becomes explicit that Soysal's explanation tacitly shifts the weight of the answer to the why-question onto the implicit assumption that the iterative conception is correct. But regardless of whether there are any strong arguments for this position, the minimalist is faced with an acute problem: her

[16] One can also turn the Collapse Problem upside down and take the set-theoretic paradoxes as evidence to support the view that Unrestricted Comprehension is not actually as natural as many people think. This view is implicit in Boolos's account of plural logic:

> Abandon [in the face of the paradoxes], if one ever had it, the idea that use of plural forms must always be understood to commit one to the existence of sets (or "classes," "collections," or "totalities") of those things to which the corresponding singular forms apply. The idea is untenable in general in any event: There are some sets of which every set that is not a member of itself is one, but there is no set of which every set that is not a member of itself is a member [...] The claim that this sentence entails the existence of this large set strikes me as most implausible: there may be a set containing all trucks, but that there is certainly doesn't seem to follow from the truth of "There are some trucks of which every truck is one." (Boolos [15], p. 442,7)

This view lead Boolos to adopt a kind of minimalism:

> But although we may guess at Frege's trains of thought, I think we cannot explain how the serpent [of paradox] entered Eden except to say: *it is a brute fact that you cannot inject the power set of a set into that set.* (Boolos [18], p. 230, emphasis is mine.)

A similar view is defended by Cartwright [23].

It seems to me, however, that the salient point is not that the occurrence of a plural in a true sentence suggests by itself that it can be sigularized, but rather that singularization is possible whenever said plural is intelligible in the sentence. And this brings us back to the Collapse Problem.

See also Linnebo [67] for a general critique of Boolos's claim that plural logic is ontologically innocent.

arguments for the correctness of the iterative conception are now implicitly incorporated in her answer, and so are not allowed to depend on any assumption that is not part of what is minimally required to understand the iterative conception.[17] Hence, proponents of the conception-based explanation must rely on an equally minimalist account of the correctness of the iterative conception. Otherwise, non-minimalist assumptions in the correctness argument will seep through to the explanation itself.

To illustrate this point, suppose, for example, that some minimalist's reason for regarding the iterative conception as correct is that it derives from some attributes of God and the Creation. In this case, the derived conception-based explanation is ultimately theological, not minimal. A discussion with such a theologically inspired minimalist will unfold along the following lines:

> **Skeptic**: "Why is V not a set according to you?"
>
> **Theological Minimalist**: "Well, since it is incompatible with Foundation and Pair, and these axioms are sanctioned by the iterative conception, which is itself correct."
>
> **Skeptic**: "But why do you think that this is indeed the correct conception of set?"
>
> **Theological Minimalist**: "Because God had to create some sets first, and then sets of those, and then sets of *those*, etc., when He created all of them. And this gives rise to the cumulative hierarchy."

It is clear that this is not a minimalist answer to the why-question. To resuscitate Soysal's argument we therefore also need to have a minimalist account of correctness. Fortunately, such an account was recently proposed by Incurvati [53]. In the next section, we shall therefore investigate whether we can safely base Soysal's explanation on Incurvati's minimalist account of correctness. However, in Section 4.5, I will argue that this attempt leads to *explanatory instability*.

[17] Incurvati [53] also runs into a similar problem. He writes:

> [the minimalist account of the iterative conception] offers a picture of the set-theoretic universe which fits well with a highly plausible diagnosis of the set-theoretic paradoxes and explains the non-existence of, e.g. the universal set and the Russell set; (Incurvati [53], p. 83)

But a diagnosis that paradox in the naive conception of set originates from a combination of open-endedness and universality (as Incurvati alleges) is not an explanation for *why* universality must be rejected so that V is not a set, unless we have already argued that the iterative conception is correct.

4.4 Inference to the Best Conception

Rather than making dubious argument on grounds of metaphysical dependency or constructibility that the iterative sets are *all* the sets there are (Subsection 3.2.1), Incurvati proposes in [53] and [54] (chapter 2) an "indirect strategy" of *inference to the best conception.*[18] He argues that the iterative conception performs better than its rivals on a number of relevant metrics. In particular, the appeal of the iterative conception is due to "the fact that it provides a convincing explanation of the paradoxes whilst showing how (most of) the axioms of standard set theory can be true."[19] Incurvati further explains that

> The minimalist does not want to say that the conception is correct only because it has these two virtues. [...] What the minimalist wants to say, rather, is that these and perhaps other virtues of the iterative conception make it the most satisfactory conception of set that we have, *and for this reason the conception should be regarded as correct.* (Incurvati [53], p. 85, emphasis is mine.)

We have argued in the previous section that the conception-based explanation inherits the underlying assumptions of our account of the correctness of the iterative conception. Now the questions is: what are those assumptions on Incurvati's account?

With regard to the virtue of justifying the axioms of standard set theory, we can lean on Potter [89] for the heavy-lifting. Without relying on any particularly contentious background assumption, he argues that the set theory **SP** (from Scott-Potter) can be derived from the iterative conception. He then shows that this theory is slightly stronger than **Z** and can formalize (most) classical mathematics.[20]

As to the second virtue, that the iterative conception does not run into the paradoxes of naive set theory, Incurvati observes that:

> We have a neat diagnosis of the paradoxes: they arise because, on the naive conception, we have both *open-endedness* and *universality* [...] It is easy to see that the cumulative hierarchy picture of the set-theoretic universe captures the set concept's open-endedness: given a

[18] See also Section 2.2.

[19] Incurvati [53], p. 84. If we want to avoid circularity here, we should be satisfied with arguing that the iterative conception does not run into the paradoxes of naive set theory. Otherwise, if we also demand that it *explains* them, the answer to the why-question becomes part of the argument for correctness, which is in turn implicitly incorporated in Soysal's proposed answer to this very same why-question. See also Footnote 17.

[20] See also Incurvati [54], pp. 48-51.

> set a occurring at a certain level of the hierarchy, there will always be a subsequent level where the larger set obtained by performing the set-formation operations on a occurs. The picture, moreover, explains the rejection of universality: the sets occurring at each level have only members of preceding levels; thus, no set can be a member of itself, which ensures that there can be no set of all sets. (Incurvati [53], pp. 70,2, emphasis as in the original.)

This explanation likewise does not seem to incur any non-minimalist debts. One might object that the rejection of universality here requires prior justification, but this critique falls flat. If we can argue by inference to the best conception that the iterative conception is correct, then this also explains why the sets are open-ended and why paradox occurs on the naive view. The point here is just that the naive conception incorporates a tension which is brought forth by the conjunction of open-endedness and universality, and that by rejecting one of them we can free this tension and thereby presumably restore consistency. The fact that the iterative conception rejects universality grants this virtue without the need for further explanation.

4.5 Soysal's Minimalism is Explanatorily Unstable

So it seems we are in good shape, inference to the best conception gives us a minimalist argument for the correctness of the iterative conception, and hence Soysal's answer to the why-question is not burdened beyond what it can carry. But which conception of set is the best? Even if we are persuaded by the arguments that Incurvati brings forth in favor of the iterative conception, this still only identifies a *family* of set conceptions that are based on iteration. Forster [34] makes precisely this point:

> The phrase 'The iterative conception of sets' conjures up a picture of a particular set-theoretic universe—the cumulative hierarchy—and the constant conjunction of phrase-with-picture is so reliable that people tend to think that the cumulative hierarchy is all there is to the iterative conception of sets: if you conceive sets iteratively, then the result is the cumulative hierarchy. [...] this is a mistake: the iterative conception of set is a good one, for all the usual reasons. However, the cumulative hierarchy is merely one way among many of working out this conception, and arguments in favour of an iterative conception have been mistaken for arguments in favour of this one special instance of it. (Forster [34], p. 97)

This caveat does not invalidate inference to the best conception, but it does show that we must be careful when we apply such reasoning to explicate. We must ensure that our explanations do not rely on spurious structural assumptions resulting from casting too wide a net when carrying out inferences of this kind.

Consider the following toy problem: what color are the sets? Imagine that we have come up with a conception of set which incorporates the standard iterative conception, and adds to it that the sets are blue. If for some reason we come to find that blueness is a plausible property of sets, we could be tempted to conclude, by inference to the best conception, that sets are indeed blue. But what weight would this argument carry? The blueness of sets does not participate in 'the iterative conception + sets are blue' being the best conception of set.

In this very contrived case the complete incidentalness of the sets being blue is easy to spot. However, the methodological lesson is that we should ask how *stable* any inference to the best conception argument is under small perturbations, such as considering the sets to be red as opposed to blue. Since the inference goes through for red sets just as well—as there seems to be no mathematico-philosophical ground to prefer one color over the other in this case—we would then have two explanations that contradict each other on the color of sets. That is, we need to be aware of and test for *explanatory instability*. Of course, this is just a toy example to illustrate a point, but the issue is not only play as I will argue now.

Forster, in the same paper [34], presents an iteration-based conception of set where V *is* a set. He asks us to imagine a "two-wands construction" of the sets: The normal iterative metaphor is that we "throw out the lasso" to circumscribe some sets and then "perform some magic on the lasso contents" with a set-forming wand to turn that plurality into a set. So, the wand functions as a kind of constructor. Forster's proposal is to add another constructor (another wand) that forms the complements of the lassoed pluralities. He then shows that this supports a theory which is mutually interpretable with **ZF** (assuming that the one-wand constructions sanctions **ZF**). With regard to whether this result legitimizes objects like V, $\{\emptyset, V\}$, and $V \setminus \omega$ as sets, or rather shows that they are a pointless epiphenomenon, Forster remarks:

> The reason why we should extend a warm welcome to theories like [this] is that they strive very hard to do something **ZF** does not even consider attempting. They represent an endeavour to reason seriously and logically about the genuinely infinite, The Unbounded. Attempts to grasp the nettle and take The Unbounded seriously (one thinks of Cantor and Dedekind) have in the past been attended with success and

> have engendered some useful new mathematics. It might be the same again. (Forster [34], p. 109)

So, we see that inference to the best conception can introduce explanatory instability when put to work to answer why-questions: there are variations of the iterative conception of set that maintain all its virtuous aspects while providing diametrically opposed answers to Soysal's why-question. The anti-minimalist can cite this as evidence that well-foundedness in its usual form is not essential to the success of the iterative conception, and that this success perhaps rather derives from recursiveness more broadly construed. This leaves the well-foundedness of the universe as an explanation for the why-question unsubstantiated, as a fish accidentally caught in too wide a net.

One might object that the one-wand construction is simpler than the two-wands construction, and that simplicity is also a virtue that should be respected in inference to the conception. This is how Incurvati deals with Forster's aberrant conception:[21]

> [Forster's theory seems] to be backed by a conception which provides us with an intuitive model [...] [This] conception has many attractive features but appears less simple than the iterative conception and disjunctive in character without leading to a theory with substantial mathematical advantages over iterative set theory from a foundational point of view. *The iterative conception, therefore, seems to fare better than this conception.* (Incurvati [54], footnote 21 on p. 183, emphasis is mine.)

Surely, however, whether a conception of set marks the right things as sets is more important than mere simplicity. Therefore, to properly contrast Forster's two-wands conception with the standard iterative conception, we would already need to know whether or not V is a set. But besides this point, the situation is untenable.

Menzel [79] argues, on the basis of classical work of Benacerraf [7], that the ordinals are not sets but rather urelements, so that there *is* a set of all of them. With this observation in mind, he argues for an adaption of the iterative conception that places the ordinals at level 0 of the cumulative hierarchy. To prevent Burali-Forti (Paradox 1), Menzel introduces a distinction between determinable and indeterminable sets (reminiscent of Cantor's view which we discussed in Subsection 1.2.1).

[21] But note that Incurvati here is not concerned with the why-question or, what I've called, explanatory instability.

The resultant theory can be proved consistent in $\ulcorner \mathbf{ZFC} + \exists\kappa(\kappa \text{ inaccessible})\urcorner$ and proves the consistency of **ZFC**.

How should we judge the relative decrease in simplicity of the conception against this urelement argument? Or against the small increase in theoretical power? How desirable is such fickleness in explanation anyway? I take it that in order to count as a good response to the paradoxes and their associated why- and what-questions, explanations should not be volatile in this sense. The problem ultimately circles back to the issue of depth, which we discussed in Chapter 3. A deep explanation that tracks 'a feature of the world' which prevents V or Ord from forming a set should not be disturbed by such conceptual perturbations.

Linnebo's explanation (Section 3.2), on the other hand, does give us the desired stability. The potentialist can rebuke Forster because the complementation-wand in the two-wands construction violates the potentiality of V.[22] And against Menzel he can maintain that regardless of whether the ordinals are urelements, since they coincide with the levels of the incompletable universe, they are not a finished totality and, therefore, cannot form a set. Thus, in contrast to Soysal's conception-based explanation, potentialism provides a stable analysis which tracks a specific feature of the world (potentiality) that not only explains *why* V is not a set, but also clarifies *what* it is instead.

4.6 Conclusion

We have seen that minimalism about the iterative conception challenges the explanatory power of potentialism since its parsimony threatens to reduce the potentialist's modal apparatus to a mere idle wheel. The question is, however, whether minimalism is able to provide a deep explanation in the same way that potentialism

[22] Forster acknowledges this issue ([34], pp. 105-6) and provides two counterarguments: 1) that the intensional nature of V in his construction is mimicked by ordinary mathematical objects such as $\mathbb{R}$; and 2) that

> surely—if [the argument that you cannot wait until the end of time] has any force at all—it has more force than [merely to discount V, $\{\emptyset, V\}$, etc. as sets]. It is not just an argument against the idea that these things are *sets*; it is an argument against the idea that they are legitimate mathematical objects *at all*. However, they surely are legitimate mathematical objects in virtue of Conway's principle [that objects may be created from earlier objects in any reasonably constructive fashion.] (Emphasis as in the original.)

As for (2), the potentialist can simply deny that complementation is a "reasonably constructive" operation. And against (1), he can argue that this view goes against the iterative view of sets in general.

can. Yablo [109], Studd [102], and Dummett ([28], chapter 24) think not.

I have interpreted the first two to resist minimalism on the ground that it cannot, from an extensional point of view, disentangle what it is about the members of some pathological set like V which conflicts with the intuition we have in our naive concept of set that whenever the members $a, b, c, \ldots$ of a plurality are determinate and exist, those members should be able to form the set $\{a, b, c, \ldots\}$. This interpretation goes beyond Soysal's rendering of their critique as being about the distinction between conceptual versus logical (syntactical) explanations. I then argued that her interpretation runs into difficulties because it cannot make this distinction in the case of the Russell paradox (which does not depend on any set-theoretical axiom). Moreover, I argued that her response provides no information about the what-question: what is V if not a set?

One can distinguish two general options in dealing with the what-question, and both lead to distinct approaches to the why-question. I showed that the first option steers us either to classes (set-like non-sets) or introduces conceptual instability. As these are both unattractive features of an account of set, I rejected Option 1 and focused attention instead on Option 2, as Soysal also seems to intend. This option led to the reformulated conception-based explanation.

I then argued that this explanation, in order to count as minimalistic, requires a minimalistically acceptable account of the correctness of the iterative conception, such as attempted in Incurvati [53]. We found, however, that this approach of inference to the best conception introduces explanatory instability if it is put to work in this way. The reason is that it cannot be used select between different explanations provided by equally virtuous conceptions within a family, despite that they may yield diametrically opposed answers to, for instance, the why-question. Examples of this phenomenon can be uncovered in Menzel [79] and Forster [34], who respectively claim that Ord and V *actually are sets* on their proposed variation of the iterative conception.

A deep explanation that tracks 'a feature of the world' which prevents V from forming a set should not be disturbed by such conceptual perturbations within the family of iterative conceptions of set. Soysal's minimalist explanation of the why-question fails to be deep in this regard. Potentialism, on the other hand, does not suffer from this defect, and so minimalism is not a proper replacement for it. I conclude that Linnebo's modal apparatus is not in vain and serves a purpose not recognized in Soysal's analysis.

Bibliography

[1] A. Avron. *Replacement*. FOM mailing list entry (2007): `https://cs.nyu.edu/pipermail/fom/2007-September/011909.html`.

[2] Z. Balogh. *A Small Dowker Space in ZFC*. Proceedings to the American Mathematical Society, pp. 2555-2560, August 1996.

[3] N. Barton, C. Ternullo, and G. Venturi. *On Forms of Justification in Set Theory*. Australasian Journal of Logic (Vol. 17, No. 4), pp. 158-200, 2020.

[4] N. Barton. *Indeterminateness and 'The' Universe of Sets: Multiversism, Potentialism, and Pluralism*. In M. Fitting (editor) *Selected Topics from Contemporary Logics* (Landscapes in Logic), College Publications, pp. 105–182, 2021.

[5] J. Baumgartner & A. Hajnal. *A proof (involving Martin's axiom) of a partition relation*. Fundamenta Mathematicae (Vol. 78, No. 3), pp. 193–203, 1973.

[6] J. Beall. *A Simple Approach Towards Recapturing Consistent Theories in Paraconsistent Settings*. The Review of Symbolic Logic (Vol. 6, No. 4), pp. 755-764, August 2013.

[7] P. Benacerraf. *What Numbers Could not Be*. The Philosophical Review (Vol. 74, No. 1), pp. 47-73, 1965.

[8] P. Benacerraf and H. Putnam (editors). *Philosophy of Mathematics*. Cambridge University Press (2nd ed.), 1984.

[9] P. Bernays. *Platonism in mathematics*. Presented as a lecture in 1935. Reprinted in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics* [8], pp. 258-71.

[10] P. Blanchette. *Frege on Consistency and Conceptual Analysis*. Philosophia Mathematica (III) (Vol. 15, No. 3), pp. 321-346, 2007.

[11] E. Boman. *Ghosts of Departed Errors: The Early Criticisms of the Calculus of Newton and Leibniz*. Juniata Voices (Vol. 17), p. 56-69, 2017.

[12] G. Boolos. *The Iterative Conception of Set*. The Journal of Philosophy (Vol. 68, No. 8), pp. 215-231, 1971.

[13] G. Boolos. *Iteration again*. Philosophical Topics XVII, no. 2, pp. 5–21, 1989.

[14] G. Boolos. *Must We Believe in Set Theory?* Unpublished, but appears in author's collected [19], pp. 120-132.

[15] G. Boolos. *To be is to be a value of a variable*. The journal of philosophy (Vol. 81), pp. 430–449, 1984.

[16] G. Boolos. *Nominalist Platonism*. The Philosophical Review (Vol. 94, No. 3), pp. 327-344, 1985.

[17] G. Boolos. *Reply to Charles Parsons' "Sets and Classes"*. Unpublished, but appears in author's [19], pp. 30-36.

[18] G. Boolos & P. Clark. *Basic Law (V)*. Proceedings of the Aristotelian Society, Supplementary Volumes (Vol. 67), pp. 213-233+235-249, 1993.

[19] G. Boolos. *Logic, Logic, and Logic*. Harvard University Press, 1998.

[20] E. Brauer. *The Modal Logic of Potential Infinity: Branching Versus Convergent Possibilities*. Erkenntnis (Vol. 87, No. 5), pp. 2161-2179, 2022.

[21] R. Cameron. *Turtles all the Way Down: Regress, Priority and Fundamentality*. Philosophical Quarterly (Vol. 58, No. 230) pp. 1-14, 2008.

[22] H. Cappelen. *Fixing Language: An Essay on Conceptual Engineering*. Oxford University Press, 2018.

[23] R. Cartwright. *Speaking of Everything*. Noûs (Vol. 28, No. 1) pp. 1-20, 1994.

[24] J. Coffa. *The Semantic Tradition from Kant to Carnap*. Cambridge University Press, 1991.

[25] M. Colyvan. *An Introduction to the Philosophy of Mathematics*. Cambridge University Press, 2012.

[26] W. Dean. *On Consistency and Existence in Mathematics*. Meeting of the Aristotelian Society held online on 15 June 2020. Proceedings of the Aristotelian Society (Vol. CXX, Part 3), 2021.

[27] W. Dean. *Hilbert on Consistency as a Guide to Mathematical Reality*. Logique Et Analyse (Vol. 237), pp. 107-128, 2017.

[28] M. Dummett. *Frege: Philosophy of Mathematics*. Harvard University Press, 1991.

[29] S. Feferman. *Predicativity*. `https://math.stanford.edu/~feferman/papers/predicativity.pdf`.

[30] S. Feferman. *Gödel's program for new axioms: Why, where, how and what?* `https://math.stanford.edu/~feferman/papers/Godelprogram.pdf`.

[31] J. Ferreirós. *The Early Development of Set Theory*, In The Stanford Encyclopedia of Philosophy (Summer 2023 Edition), E. Zalta & U. Nodelman (eds.), `https://plato.stanford.edu/archives/sum2023/entries/settheory-early/`.

[32] E. Ferrier. *Against the iterative conception of set*. Philosophical Studies (Vol. 176, No. 10), pp. 2681-2703, 2019.

[33] T. Forster. *The Axioms of Set Theory*. `https://www.dpmms.cam.ac.uk/~tf/axiomsofsettheory.pdf`.

[34] T. Forster. *The Iterative Conception of Set*. The Review of Symbolic Logic (Vol. 1, No. 1), pp. 97-110, June 2008.

[35] B. van Fraassen. *The Scientific Image*. Clarendon Press, 1980.

[36] M. Frapolli. *Is Cantorian set theory an iterative conception of set?* Modern Logic (Vol. 1, No. 4), pp. 302-318, Spring 1991.

[37] H. Friedman. Some posts from a debate between H. Friedman and C. McLarty on the the FOM mailing list (1999): `https://cs.nyu.edu/pipermail/fom/1999-April/002983.html`, `https://cs.nyu.edu/pipermail/fom/1999-April/003014.html`.

[38] H. Friedman. FOM mailing list (2007): `https://cs.nyu.edu/pipermail/fom/2007-August/011804.html`.

[39] K. Gödel. *What is Cantor's Continuum Problem?* The American Mathematical Monthly, Vol. 54, No. 9, pp. 515-525, 1947.

[40] M. Hallett. *Cantorian Set Theory and the Limitation of Size*. Clarendon Press, 1984.

[41] J. Hamkins. *Is the Dream Solution of the Continuum Hypothesis Attainable?*. Notre Dame Journal of Formal Logic (Vol. 56, No. 1), pp. 135-145, 2015.

[42] J. Hamkins. *The Set-Theoretic Multiverse*. The Review of Symbolic Logic (Vol. 5, No. 3), pp. 416-449, 2012.

[43] G. Hardy. *A Mathematician's Apology*. Cambridge University Press, Canto edition, 1992 (first published in 1940).

[44] W. Hart. *The Potential Infinite*. Proceedings of the Aristotelian Society (Vol. 76), pp. 247-264, 1976.

[45] T. Heath. *The Thirteen Books of the Elements*. Original publication: *Translation of Euclid's Elements from the text of Heiberg with introduction and commentary*, Cambridge University Press, 1908 (2nd ed. 1925). Reprinted by Dover in 1956 under the listed title.

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

[47] R. Hersh. *Mathematics has a Front and a Back*. Synthese (Vol. 88, No. 2), pp. 127-133, 1991 (first published in 1988).

[48] D. Hilbert. *On the Infinite*. Delivered as lecture in 1925. Reprinted in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics* [8], pp. 183-201.

[49] I. Hacking. *Why is there Philosophy of Mathematics at all?* Cambridge University Press, 2014.

[50] J. Hamkins. *The Modal Logic of Arithmetic Potentialism and the Universal Algorithm*. Pre-print: `https://arxiv.org/abs/1801.04599v3`.

[51] J. Hamkins. *Kuratowski's definition of ordered pairs*. Answer on MathOverflow (2011): `https://mathoverflow.net/q/62801`.

[52] J. Hamkins & Ø. Linnebo. *The Modal Logic of Set-Theoretic Potentialism and the Potentialist Maximality Principles*. The Review of Symbolic Logic (Vol. 15, No. 1), pp. 1-35, 2022.

[53] L. Incurvati. *How to be a minimalist about sets*. Philosophical Studies, 159:69–87, 2012.

[54] L. Incurvati. *Conceptions of Set and the Foundations of Mathematics*. Cambridge University Press, 2020.

[55] D. Isaacson. *The Reality of Mathematics and the Case of Set Theory*. In Z. Novak & A. Simonyi (eds.), *Truth, Reference and Realism*, Central European University Press, pp. 1-76, 2010.

[56] H. Keisler. *Ultraproducts and elementary models*. Indagationes Mathematicae (Vol. 23), pp. 477–495, 1961.

[57] K. Kunen. *Set Theory*. College Publications, 2011.

[58] I. Jané. *The Role of the Absolute Infinite in Cantor's Conception of Set*. Erkenntnis (Vol. 42, No. 3), pp. 375-402, May 1995.

[59] A. Kanamori. *In Praise of Replacement*. The Bulletin of Symbolic Logic, pp. 46-90, March 2012.

[60] P. Koellner. *Large Cardinals and Determinacy*. In the Stanford Encyclopedia of Philosophy, `https://plato.stanford.edu/entries/large-cardinals-determinacy/`.

[61] P. Koellner. *On the question of absolute undecidability*. In S. Feferman, C. Parsons & S. Simpson (eds.), *Kurt Gödel*, Association for Symbolic Logic, pp. 153-188, 2010.

[62] I. Lakatos. *Infinite regress and foundations of mathematics*. In J. Worrall & G. Currie (eds.), *Mathematics, Science and Epistemology*, Cambridge University Press, pp. 3-23, 1978 (the original publication in from 1962).

[63] I. Lakatos. *A Renaissance of Empiricism in the Recent Philosophy of Mathematics*. The British Journal for the Philosophy of Science (Vol. 27, No. 3), pp. 201-223, 1976.

[64] I. Lakatos. *Proofs and Refutations*. Cambridge University Press, Cambridge Philosophy Classics edition of 2015 (first published in 1976).

[65] S. Lavine. *Understanding the Infinite*. Harvard University Press, 1998.

[66] T. Leinster. *Borel Determinacy Does Not Require Replacement*. The *n*-Category Café blog (2021): `https://golem.ph.utexas.edu/category/2021/07/borel_determinacy_does_not_require_replacement.html`.

[67] Ø. Linnebo. *Plural Quantification Exposed*. Noûs (Vol. 37, No. 1), pp. 71-92, 2003.

[68] Ø. Linnebo. *The Potential Hierarchy of Sets*. The Review of Symbolic Logic (Vol. 6, No. 2), pp. 205-228, 2013.

[69] Ø. Linnebo. *Pluralities and Sets*. The Journal of Philosophy (Vol. 107, No. 3), p. 144-164, 2010.

[70] Ø. Linnebo. *Putnam on Mathematics as Modal Logic*. Chapter 14 in G. Hellman & R. Cook (eds.), *Hilary Putnam on Logic and Mathematics*, Springer Verlag, 2018.

[71] Ø. Linnebo. *Plurals and Modals*. Canadian Journal of Philosophy (Vol. 46, Nos. 4-5), pp. 654-676, 2016.

[72] Ø. Linnebo & S. Shapiro. *Actual and Potential Infinity*. Noûs (Vol. 53, No. 1), pp. 160–191, 2019.

[73] W. Lycan. *Explanationist Rebuttals (Coherentism Defended Again)*. The Southern Journal of Philosophy (Vol. 50, No. 2), pp. 5-20, 2012.

[74] P. Maddy. *Believing the Axioms I*. The Journal of Symbolic Logic (Vol. 53, No. 2), pp. 481-511, 1988.

[75] P. Maddy. *Some Naturalistic Reflections on Set Theoretic Method*. Topoi (Vol. 20, No. 1), pp. 17-27, 2001.

[76] P. Maddy. *Set-Theoretic Foundations*. Contemporary Mathematics (Vol. 690), 2017.

[77] P. Maddy. *What Do We Want a Foundation to Do?* Chapter 13 in S. Centrone et al. (eds.), *Reflections on the Foundations of Mathematics*, Springer (Synthese Library Volume 407), 2019.

[78] M. Magidor. *Some Set Theories are More Equal*. Unpublished EFI paper (`https://logic.harvard.edu/efi/`).

[79] C. Menzel. *Cantor and the Burali-Forti Paradox*. The Monist, Vol. 67, No. 1, pp. 92-107, January 1984.

[80] C. Menzel. *Wide Sets, ZFCU, and the iterative conception*. The Journal of Philosophy, Vol. 111, No. 2, pp. 57-83, 2014.

[81] P. Murphy. *Coherentism in Epistemology*. In the Internet Encyclopedia of Philosophy, `https://iep.utm.edu/coherentism-in-epistemology/`.

[82] E. Nelson. *Warning Signs of a Possible Collapse of Contemporary Mathematics*. In: Infinity. Cambridge Univ. Press, pp. 76–85, 2011.

[83] A. Oliver and T. Smiley. *What are sets and what are they for?* Philosophical Perspectives (Vol. 20, No. 1), pp. 123–155, 2006.

[84] C. Parsons. *Sets and Classes*. Noûs (Vol. 8, No. 1), pp. 1-12, March 1974.

[85] C. Parsons. *What is the Iterative Conception of Set?* Originally published in: R. Butts, J. Hintikka (eds.) *Logic, Foundations of Mathematics, and Computability Theory* (1977). Reprinted in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics* [8], pp. 503-29.

[86] A. Paseau. *Boolos on the Justification of Set Theory*. Philosophia Mathematica (Vol. 15, No. 1), 2007.

[87] A. Pease, A. Aberdein, and U. Martin. *Explanation in mathematical conversations: An empirical investigation*. Philosophical Transactions of the Royal Society A (377), 2019.

[88] T. Poston. *Reason and Explanation: A Defense of Explanatory Coherentism*. Palgrave Macmillan, 2014.

[89] M. Potter. *Set Theory and its Philosophy*. Oxford University Press, 2004.

[90] M. Potter. *Iterative Set Theory*. The Philosophical Quarterly (Vol. 43, No. 171), pp. 178-193, 1993.

[91] M. Potter. *Intuitive and Regressive Justifications*. Philosophia Mathematica (Vol. 28, No. 3), pp. 385–394, 2020.

[92] G. Priest. *In Contradiction*. Oxford University Press, 2006 (second edition).

[93] H. Putnam. *Mathematics without foundations*. Journal of Philosophy (Vol. 64, No. 1), pp. 5-22, 1967.

[94] W. Quine. *Two dogmas of empiricism*. Philosophical Review 60, pp. 20-43, 1951.

[95] D. Schlimm. *Mathematical Concepts and Investigative Practice*. In U. Feest & F. Steinle (eds.), *Scientific Concepts and Investigative Practice*, De Gruyter, pp. 127-148, 2012.

[96] J. Serre. *How to use finite fields for problems concerning infinite fields*. `https://arxiv.org/abs/0903.0517`, 2009.

[97] S. Shapiro. *Computability, Proof, and Open-Texture*. In A. Olszweski, J. Wolénski & R. Janusz (eds.), *Church's Thesis After 70 Years*, De Gruyter, pp. 420–451, 2006.

[98] S. Shelah. *Every two elementarily equivalent models have isomorphic ultrapowers*. Israel Journal of Mathematics (Vol. 10), pp. 224–233, 1971.

[99] D. Sherry. *The Wake of Berkeley's Analyst: Rigor Mathematicae?* Studies in History and Philosophy of Science (Vol. 18, No. 4), pp. 455-480, 1987.

[100] Z. Soysal. *Why is the universe of sets not a set?* Synthese (Vol. 197, No. 2), pp. 575–597, 2020.

[101] J. Steel. *Gödel's program*. In J. Kennedy (ed.), *Interpreting Gödel*, Cambridge University Press, pp. 153-179, 2014.

[102] J. Studd. *The Iterative Conception of Set: A (Bi-)Modal Axiomatisation*. Journal of Philosophical Logic, 42:697–725, 2013.

[103] W. Tait. *Constructing Cardinals from Below*. `http://home.uchicago.edu/~wwtx/Foundations3.pdf`.

[104] W. Tait. *Finitism*. The Journal of Philosophy (Vol. 78, No. 9), pp. 524-546, 1981.

[105] G. Takeuti. *Two Applications of Logic to Mathematics*. Originally published in 1978, reprinted by Princeton University Press, 2015.

[106] F. Tanswell. *Conceptual engineering for mathematical concepts*. Inquiry (Vol. 61, No. 8), pp. 881–913, 2018.

[107] C. Ternullo. *Maddy On The Multiverse*. Chapter 3 in S. Centrone, D. Kant & D. Sarikaya (eds.), *Reflections on the Foundations of Mathematics: Univalent Foundations, Set Theory and General Thoughts*, Springer Verlag, pp. 43-78, 2019.

[108] P. Thagard. *Explanatory Coherence*. Behavioral and Brain Sciences (Vol. 12, No. 3), pp. 435-467, 1989.

[109] S. Yablo. *Circularity and paradox*. In V. F. H. Thomas Bolander & S. A. Pedersen (Eds.), *Selfreference*, 2004.

[110] F. Waismann. *Verifiability*. In *Logic and Language* (A. Flew ed.), Basil Blackwell, 1968.

[111] D. Waxman. *Imagining the Infinite*. Draft version: `https://www.danielwaxman.com/files/infinite.pdf`.

[112] D. Waxman. *Mathematical Freedom, Truth, and Coherence*. Draft version: `https://www.danielwaxman.com/files/freedom.pdf`.

[113] L. Wittgenstein. *Remarks on the Foundations of Mathematics*. G. von Wright, R. Rhees, and G. Anscombe (eds.). Translated by G. Anscombe. Blackwell, 3rd revised edition, 1978.

[114] M. Zelcer. *Against Mathematical Explanation*. Journal for General Philosophy of Science (Vol. 44, No. 1), pp. 173-192, 2013.