Formal Analyticity

Formal Analyticity

A dissertation presented
by
Zeynep Soysal
to
The Department of Philosophy

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Philosophy

Harvard University
Cambridge, Massachusetts
May 2017

Dissertation Advisor: Professor Peter Koellner Zeynep Soysal

Formal Analyticity

Abstract

This dissertation consists of three papers that together serve to defend a notion of analyticity for formal languages: A sentence (or rule of inference) of a formal language is *formally analytic* if understanding the sentence (or rule of inference) is sufficient for being in a position to know it. This is a way of reviving a traditional method of explaining our knowledge of mathematics. I first defend the claim that certain basic axioms of set theory are formally analytic by using a notion of unfolding, inspired by Kurt Gödel's influential remarks that some axioms of set theory "only unfold the content of the concept of set" (in "Unfolding the Content of the Concept of Set"). I then argue that Timothy Williamson's famous argument template against analyticity doesn't work against the claim that there are formally analytic sentences (in "Defending Formal Analyticity"). Finally, I argue that recognizing formal analyticity in set theory can help answer some philosophical questions concerning the nature and extent of the universe of sets (in "Why Is the Universe of Sets Not a Set?").

Contents

Annem ve babam için

Acknowledgments

So many people contributed to the completion of this dissertation. It is a great pleasure to look back and think about them all, and to thank them here.

First of all, I would like to thank my committee members, Peter Koellner, Bernhard Nickel, Charles Parsons, and Mark Richard, for all of their teaching, guidance, support, encouragement, and kindness. I am incredibly grateful for having had the opportunity to learn everything from you.

For getting me started, and still to this day providing me valuable guidance and support, I would like to thank Matti Eklund and Harold Hodes.

For their comments, discussions, and unwavering support throughout my graduate studies, I would like to thank the faculty members of the Harvard Philosophy Department, and, in particular, Selim Berker, Warren Goldfarb, Ned Hall, Rusty Jones, Jeffrey McDonough, Dick Moran, Susanna Rinard, Jacob Rosen, Susanna Siegel, Alison Simmons, and Hugh Woodin.

I would like to thank Jeffrey McDonough also for sparking my interest in Leibniz, and driving me to do work on Leibniz that, even though it didn't make it in the text of this dissertation, has been an important and dear component of my graduate work.

Without my friends and colleagues, nothing good would have been possible. I am incredibly grateful in particular to Rachel Achs, Diana Acosta Navas, Arden Ali, Efe Balıkçıoğlu, Neil Barton, Ewa Bigaj, Douglas Blue, James Bondarchuk, Ekaterina Botchkina, Ryan Davis, Sandy Diehl, Noel Dominguez, Marc Gasser-Wingate, Gabriel Goldberg, Ronni Gura Sadovsky, Ezgi Hacısuleyman, Bill Hoston, Lauren Kopajtic, Cristina Malatesta, Paul Marcucilli, Patricia Marechal, Jack Marley-Payne, Eylem Özaltun, Thomas Pendlebury, Alexander Prescott-Couch, Michael Rabenberg, Shantia Rahimian, Richard Sanchez, Aleksy Tarasenko-Struc, Mary-Jane Tsang Mui Ching, Kate Vrendenburgh, Davide and Michele Zaccagnini.

Olivia Bailey, Lauren Davidson, Doug Kremm, Elizabeth Miller, and Rebecca Millsop: I simply cannot thank you enough.

I would like to thank my dear friends Caroline Corn and Charlotte Kreitmann, for their friendship from long before I started this adventure to the very end of it.

I would like to thank my parents, for being the most optimistic, kind, generous, encouraging, supportive, and wise people I know.

Finally, I would like to thank Jens Kipper, for support, for philosophy, and most importantly, for happiness.

0

Introduction

At the highest level of generality, the question that has been guiding this dissertation is: How are we justified in believing axioms and rules of inference of mathematics? Axioms and rules of inference are "first principles" in mathematics—they are principles that are accepted and relied on by mathematicians without proof. My guiding question is perhaps the most central foundational question in the epistemology of mathematics: How are we justified in believing the first principles of mathematics?

The answer to this question that I have always found most appealing is that mathematical axioms and rules of inference are *analytic* or *conceptually true*: They are true only because of the meanings of the words in them, and whoever understands the meanings of those words is thereby justified in believing these axioms and rules of inference. This answer was made popular by the logical empiricists in the early twentieth century, but it then quickly lost traction, as philosophers such as Willard van Orman Quine and Hilary Putnam raised important and general problems for the idea that there are any analytic or conceptual truths. In particular, these philosophers challenged the idea that conceptual or linguistic competence has certain "fixed-points," that is, that there is a sharp distinction between what is constitutive of meaning or linguistic competence—the fixed-points—and what is not.

This dissertation aims to revive the logical empiricists' answer to the guiding question by defending the idea that there is a perfectly sensible and importantly useful notion of analyticity—formal analyticity—according to which mathematical axioms and rules of inference are analytic. More specifically, I aim to show that there are fixed-points for conceptual and linguistic understanding in formal mathematical languages. My arguments are based on the observation that formal mathematical concepts and expressions work very differently from concepts and expressions of ordinary language, both from the point of view of the practice surrounding those concepts and expressions—the practice of teaching, using, and ascribing understanding of those concepts and expressions—and from the point of view of their metasemantics. I argue that it is because of this peculiar nature of mathematical concepts and expressions that we can defend a notion of analyticity for mathematical languages, even if one perhaps cannot maintain that there are analytic sentences in natural language. Finally, I show the utility of recognizing formal analyticity for making progress on certain important questions in the philosophy of mathematics.

These are the broad strokes of the questions and challenges that have guided the writing of this dissertation. Let me now introduce some elements of my argumentative strategy for formal analyticity.

0.1 Formal Analyticity

In this dissertation, I am primarily concerned with defending the claim that certain axioms of set theory are formally analytic. Set theory provides a particularly good case for formal analyticity, in part because of the central role that the axioms of set theory play—and have played from very early on—in the practice of doing and teaching set theory. The claim that certain axioms of set theory are analytic is also of contemporary interest: Many contemporary set theorists have been interested in Gödel's influential remarks that some axioms of set theory "only unfold the content of the concept of set," in the hope that they can use this idea to justify new axioms for set theory. As I argue in "Unfolding the Content of the Concept of Set" (Chapter 1), Gödel can plausibly be interpreted as claiming that some axioms of set theory are analytic.

In broad outline, the claim that some axiom of set theory is formally analytic is the claim that understanding the concept SET (or understanding the meanings of expressions such as 'set' or '∈') is sufficient for being in a position to know that axiom. There are two readings of what it is to "know an axiom"—one weaker, and one stronger. There are thus really two versions of my main claim that certain axioms of set theory are formally analytic. I argue for the stronger claim in Chapter 1, and for the weaker claim in "Defending Formal Analyticity" (Chapter 2). Before I explain the difference between the two claims, let me make some preliminary clarifications.

The first clarification concerns mathematical propositions. On a natural view, the objects of knowledge (and epistemic access more generally) and the objects of belief (and intentional access more generally) are *propositions*. But there is a difficulty in speaking about mathematical propositions. According to some prominent theories of propositions, mathematical propositions are too coarse-grained to be of any use in an account of the epistemology of mathematics. Possible-world semantics, for instance, takes the proposition expressed by a sentence to be the sets of all (metaphysically) possible worlds with respect to which the sentence is true. As it is widely assumed that (meaningful) mathematical sentences are either true in all possible worlds or false in all possible worlds, all true mathematical sentences, on this view, express a unique proposition: the set of all possible worlds. It follows that whoever knows one mathematical proposition knows them all. There are several ways to try to remedy this problem.

One can try to stay within the framework of possible-worlds semantics, and explain the content of mathematical knowledge as metalinguistic and contingent: On this strategy, advanced famously by Robert Stalnaker, what mathematicians really come to know are contingent truths about which mathematical sentences express the one necessary truth. Alternatively, one can try to construe propositions in a more fine-grained manner, for instance by taking them to be structured intensions or properties, or to be constituted of Fregean senses. One can also try to avoid the issue altogether by taking (meaningful) mathematical sentences to be the objects of epistemic and intentional access, even though this sounds somewhat unnatural. All these options have their costs and challenges. Throughout this dissertation, I employ a naïve view according to mathematical propositions are fine-grained enough—whatever mathematical propositions are, it should turn out, for instance, that the proposition *every non-empty set has an* $\in$*-minimal element* is distinct from the proposition *every set has a powerset.* But I don't take a further stand on how to best construe mathematical propositions. In Chapter 1 and in "Why Is the Universe of Sets Not a Set?" (Chapter 3), I simply speak as though mathematical propositions are fine-grained enough for my purposes, while in Chapter 2, I don't speak of propositions but only of sentences of formal languages. The issue of how to construe mathematical propositions doesn't affect my arguments in this dissertation, but it is an important question that needs to be investigated as part of the general project of developing an epistemology of mathematics.

The second clarification concerns concepts (and it only concerns Chapters 1 and 3).

Throughout this dissertation, I adopt what I take to be what is as close as possible to maximally neutral view of what are concepts. Let me call this here the 'Minimal View of Concepts'. According to the Minimal View of Concepts, concepts are constituents of the contents of attitudes such as beliefs, thoughts, knowledge, and so on. What it is to possess a concept (and I will always use 'possess a concept', 'grasp a concept', and 'understand a concept' interchangeably), in turn, just is to have attitudes with contents of which the concept is a constituent. Since the content of a state is generally understood to be something that determines (or something that *just is*) the state's truth (or veridicality, or satisfaction) conditions, according to the Minimal View of Concepts, a concept also determines (or perhaps even *just is*) a partition of all (actual and possible) things into two classes: the class of all (actual or possible) things that "fall under" the concept, and the rest. In Chapter 1, I call the partition determined by a concept the 'content' of the concept. This is a slightly loose way of speaking (since concepts themselves are constituents of contents according to the Minimal View of Concepts), but it fits well with Gödel's own way of talking, and with the Linguistic View of Concepts, which I adopt in Chapter 1 (and which will explain shortly). One way to think about the partition determined by a concept is as an intension: a function from possible worlds to sets of things at that world that "fall under" the concept.

One advantage of the Minimal View of Concepts is that it doesn't stack the deck against opponents of analyticity: I agree with Timothy Williamson's thought that accounts of concept possession according to which possessing a concept is a matter of having knowledge of, or (in the spirit of George Bealer's account) reliable access to what falls under the concept in effect "build the consequent into the antecedent," if what one wants to defend is precisely the existence of such links between concept possession and knowledge.

The Linguistic View of Concepts says slightly more than the Minimal View of Concepts. According to the Linguistic View of Concepts, concepts are meanings of expressions; meaningful expressions are said to "express" concepts. In particular, if a certain linguistic community uses a meaningful expression, *e*, we can say that their expression expresses some concept, C. At the level of the individual, if a person, S, uses an expression, *e*, with a certain meaning in her own idiolect, then S means C by *e* in her idiolect. So, on the Linguistic View of Concepts, to say that S possesses C just is to say that S means C by *e* (in her idiolect). Externalists (such as Tyler Burge) and internalists (such as Christopher Peacocke) alike adopt the Linguistic View of Concepts.

In Chapter 1, I also adopt the Linguistic View of Concepts. The Linguistic View of Concept merges with the Minimal View of Concepts at the level of concepts that are expressed in language: If a person, S, means C by *e* in her idiolect (so if C is expressed in S's own language), then, presumably, S has propositional attitudes with contents that contain C. Since I am only concerned with concepts that are in fact expressed in a shared mathematical language, such as the concept SET, this restriction to concepts expressed in language doesn't affect my arguments in this dissertation.

Let me make one final clarification about concepts. Whatever the correct account of concepts, it should presumably turn out that the concept POWERSET OF THE EMPTY SET is distinct from the concept SET OF THE EMPTY SET, even though they both determine the same extension in every possible world (presumably, they both pick out $\{\emptyset\}$ at every possible world). One might claim that POWERSET OF THE EMPTY SET and SET OF THE EMPTY SET are distinct concepts simply because they are complex (i.e. non-lexical) concepts constituted of distinct lexical concepts (for instance, the first "contains" the concept POWERSET while the second doesn't). This is a natural thought: After all, the expression 'powerset of the empty set' is also distinct from 'set of the empty set'. But we should be careful to not go too fine-grained in the individuation of concepts and end up having a distinct concept for each distinct linguistic expression: It should (presumably) turn out, for instance, that 'ensemble vide', 'empty set', and '$\emptyset$' all express the same concept. These are tricky issues, and it is not clear to me how best to resolve them. They do not affect my arguments in this dissertation, but I do think more work needs to be done to clarify these issues, as they may have implications for the intelligibility of speaking of (mathematical) concepts in general.

With these preliminary clarifications in place, let us return to the main claim, namely that understanding the concept SET (or understanding the meaning of expressions such as 'set' or '$\in$') is sufficient for being in a position to know certain axioms (where these axioms are the ones I call 'formally analytic').

In Chapter 1, I argue that what understanding provides is knowledge of certain conditionals of the form 'if there are sets, then A', where A an axiom of set theory; A expresses the proposition that sets have certain properties, such as the property of being extensional, or the property of being well-founded, or the property of having powersets, and so on. This is the strong sense in which understanding provides "knowledge of the axioms."

In Chapter 2, I argue that what understanding provides is knowledge that certain sentences have a privileged status within a given formal system; for instance, the status of being an axiom or rule of inference of that formal system, or the status of being usable as an assumption in derivations of the formal system. This is the weak sense in which understanding provides "knowledge of the axioms."

The strong sense is epistemically stronger: Just because a sentence is an axiom of set theory doesn't mean that sets, if they exist, are as the axiom says they are. Indeed, an externalist line of thought is that set theorists could be radically mistaken about sets; it may be that none of the axioms of set theory—none of the principles that are typically presupposed among set theorists—are true. Sets will be as the axiom says they are if, for instance, axioms of formal systems have a content-determining status. So, for instance, if the correct metasemantic account of the expressions of set theory is one according to which the fact that a sentence is accepted as an axiom in a community of mathematicians makes it the case that the sentence helps "fix the meaning" of the expressions of that formal system, then the weak claim will carry more epistemic import. I discuss this issue in Section 2.5.

Throughout this dissertation, I develop two strategies for defending my main claims that some axioms are formally analytic. The first strategy, which I shall call here 'the Strategy from Minimal Competence', starts off from an intuitively plausible general condition on conceptual and linguistic understanding. The second strategy, which I shall call here 'the Strategy from Metasemantics', starts off from metasemantic theories (also called 'foundational theories of meaning')—that is, from theories of what makes it the case that one possesses a concept (or uses an expression with a particular meaning). Although I only employ the Strategy from Minimal Competence to defend the weaker claim in Chapter 2 (for reasons that will become clear in Section 0.2), the Strategy from Metasemantics can also help the weaker claim have more epistemic import, as I explain in Section 2.5; the two strategies thus supplement each other throughout the dissertation.

0.2 The Strategy from Minimal Competence

The first strategy, which is at play in both Chapter 1 and 2, is the Strategy from Minimal Competence. It has three steps. The first step is to argue for a general condition on understanding: Understanding a concept requires having minimal competence with the concept. This condition is widely accepted; it also fits with our general practice of ascribing conceptual understanding. The second step is to argue that minimal competence with mathematical concepts just is basic mathematical competence, i.e. the competence to solve basic mathematical problems. The competence to solve basic mathematical problems may for instance include the competence to perform elementary arithmetical operations (e.g. to add and subtract), to solve basic problems of set theory, or to write down the most basic derivations in a formal system. The third and final step is to argue that having these basic mathematical competences requires (i) knowing the conditionally weakened axioms (in Chapter 1), and (ii) knowing that the axioms are axioms of the formal system in question (in Chapter 2).

In Chapter 2, my opponents accept the first step of the Strategy from Minimal Competence. Moreover, the claim that I argue for in Chapter 2, (ii), is an epistemically modest claim. The Strategy from Minimal Competence by itself is thus sufficient for the purposes of Chapter 2.

In Chapter 1, matters are slightly more complicated. Firstly, the claim I want to defend, (i), is more ambitious: I want to argue that minimal competence requires knowledge of certain conditionals about sets, not just knowledge that certain sentences are axioms within a given formal system. As a result, I work with a slightly stronger notion of "basic mathematical competence" in Chapter 1 than in Chapter 2; for instance, I argue that having the competence to solve basic mathematical problems requires gaining some knowledge (see for instance the discussion of Case 2 in Section 1.1.2). Secondly, the dialectical situation of Chapter 1 is different; there, I am not arguing against opponents who are happy to accept my claim that understanding a concept requires having minimal competence with the concept. I do think that the claim that concept possession requires minimal competence is highly plausible; it is also very widely accepted. But it is not entirely clear why it must be true from a theoretical point of view; in particular, if one holds a view according to which having a concept is merely a matter of having certain thought contents, and if one requires very little by way of abilities for having thought contents (as some externalists about content would argue), then it is not entirely clear how one would defend the claim that concept possession in general requires having minimal competence with the concept. This is the main reason why I supplement the Strategy from Minimal Competence with the Strategy from Metasemantics in Chapter 1. As I mentioned earlier, the Strategy from Metasemantics is also relevant for getting more epistemic import from the claim that axioms are formally analytic defended in Chapter 2—so the two papers, and the two strategies, are intimately connected.

0.3 The Strategy from Metasemantics

In outline, the Strategy from Metasemantics is to examine accounts of what determines that one possesses SET (or of what determines that one uses 'set' to mean SET), and then argue that (a) at least some of these accounts imply that whoever possesses SET in the way specified by these accounts is thereby in a position to know some of the axioms of set theory, and (b) the most realistic accounts of how set theorists in fact possess SET are accounts such as in (a). The key challenge for this strategy arises in dealing with externalist accounts of concept possession—accounts according to which factors that determine that one means SET by 'set' are not a priori accessible to the thinker. In my view, there are two such accounts that are worthy of serious investigation:

Deference to an Optimal Theory, according to which one possesses SET by deferring to an optimal theory (which I examine in Section 1.2.2), and *Deference to Other People*, according to which one possesses SET by deferring to others (which I examine in Section 1.2.3). In outline, my strategy here is as follows. I first argue that if one possesses a concept by deference—either to an optimal theory, or to other people—then one has certain dispositions or intentions to defer. I then argue that, realistically, most set theorists don't have the disposition or the intention to defer at least concerning the most basic axioms of set theory (in fact, one could make the stronger claim that they have the disposition *not* to defer concerning these axioms). I use these two claims to show that the way set theorists actually understand the concept SET is such that their understanding of SET puts them in a position to know at least some basic axioms of set theory.

My claims about the intentions and dispositions of set theorists can be disputed; at least, they are not supported by empirical research. I therefore also examine what would happen if these claims were false. In Section 1.2.3, I argue that, even on an extremely deferential account of the possession of SET, there is at least going to be a link between understanding and knowledge of the axioms at the level of *groups*—I argue that the practice of set theory is impossible unless the community of people who practice set theory as a whole is in a position to know the axioms of set theory. Even though this is no longer exactly the claim that certain axioms of set theory are formally analytic, it is an interesting claim that could enable us to start developing a social view of analyticity and a priority.

In outline, then, the Strategy from Metasemantics is really a way of defending the claim that certain axioms are formally analytic relative to ways of possessing concepts (or under the assumption that a certain metasemantic account it correct). If there really are different ways of possessing concepts, then it should be no surprise that the notion of formal analyticity will also have to take ways of possessing concepts as parameters.

0.4 Applications

In Chapter 3, the final chapter of this dissertation, I turn to examining one application of the claim that some axioms of set theory are formally analytic. The application is to a debate between actualists and potentialists concerning the question of why the universe of sets is not a set. On a naïve view, the explanation for why there is no set of all sets is that if there was a set of all sets, it would be a member of itself, but sets cannot be members of themselves, so there is no set of all sets—it's as simple as that. My aim in Chapter 3 is to show that, contrary to what most actualists and potentialists think, this "minimal explanation" is as good an explanation for why there is no set of all sets as we can get. I first argue against the actualists' and potentialists' alternative explanations of why there is no set of all sets: I argue that they rely on the minimal explanation and add nothing to it. I then explain why the minimal explanation is the best explanation of why there is no set of all sets.

Here is, in outline, the main idea. The minimal explanation appeals to a feature of sets: Sets are well-founded. Opponents of the minimal explanation are unhappy with this. However, the fact that sets are well-founded follows from the iterative conception of sets, and everyone in the debate—opponents and proponents of the minimal explanation alike—presuppose the iterative conception of sets: Everyone who asks the question of why there is no set of all sets presupposes that sets are arranged in the cumulative hierarchy of sets. So, I argue that at the very least because everyone in the debate presupposes the iterative conception of sets—and hence also that sets are well-founded—it is all right to appeal to the fact that sets are well-founded in explaining why there is no set of all sets.

This (together with my criticisms of the alternatives) shows that the minimal explanation is at least as good an explanation for why there is no set of all sets as the alternatives. But why is the minimal explanation the best possible explanation for why there is no set of all sets? This is where formal analyticity comes in play. In my view—in the view that started to defend in this dissertation—certain basic axioms of set theory, including Foundation, are conceptually true. I think that they are conceptually true both in the sense of being formally analytic—that is, whoever understands them is in a position to know them. But I also think they are conceptually true in the sense of being content-determining (which is something I argue for throughout the Strategy from Metasemantics): The fact that set theorists have adopted these as basic axioms of set theory helps makes it the case that the expressions of set theory mean what they do—and, in particular, that these basic axioms are true.

If this view is correct, then the minimal explanation is a kind of conceptual explanation: The universe of sets is not a set because it follows from what we mean by 'set' that it isn't. And that's as deep an explanation as we can get. Indeed, in general, being told that something is a conceptual truth, or that something follows from a conceptual truth, makes for a satisfactory explanation. Take the following, simplistic example. If I am told that it is a conceptual truth that knowledge entails truth, it would be absurd for me to keep asking "But why can't knowledge be false?" What the minimal explanation does in the case of the question of why there is no set of all sets is provide some conceptual truths about sets, and then make intelligible how the fact that the universe of sets is not a set follows from those conceptual truths. Given this explanation, it would be absurd to keep asking "But why can't the universe of sets be a set?"

I foresee formal analyticity to have many more applications, both in debates concerning mathematics, and more generally concerning the a priori. There is also more to be done to characterize and defend various aspects of formal analyticity. This dissertation is only the first step of the development of the notion of formal analyticity.

1

Unfolding the Content of the Concept of Set

In his famous paper on the continuum problem, Kurt Gödel makes the following, remarkable statements:

> [T]here may exist, beside the usual axioms, the axioms of infinity, and …[some large cardinal axioms] other (hitherto unknown) axioms of set theory which a more profound understanding of the concepts underlying logic and mathematics would enable us to recognize as implied by these concepts. (Gödel, 1964, 477)
>
> These axioms [Inaccessibles and Mahlo] show clearly not only that the axiomatic system of set theory as used today is incomplete, but that it can be supplemented without arbitrariness by new axioms which only unfold the content of the concept of *set*. (Gödel, 1964, 476f.)[1]

These statements are remarkable in part because they suggest that mere understanding of concepts is sufficient for being in a position to know some axioms of set theory.[2] This is to say that some axioms of set theory are, in some sense, analytic or conceptually true.[3] The idea that there are analytic or conceptual truths has fallen out of favor in philosophy since the time Gödel made these statements, based mainly on considerations about the viability of analyticity or conceptual truth for ordinary language or the language of natural sciences.[4] I think, however, that Gödel (or at least Gödel as I interpret him) was right, and that he was right because his claims were about mathematical concepts. Paying attention to the practice surrounding mathematical concepts—both the ordinary practice of ascribing mathematical concepts, and the practice of mathematics

[1]Of note is that in the 1947 version of this paper, Gödel uses "which are only the natural continuation of the series of those set up so far" (Gödel, 1947, 520) instead of "which only unfold the content of the concept of set." He also replaces the word "natural" in (1947) with "systematic" in (1964) when he is talking about the representation of infinite cardinal numbers.

[2]Throughout this paper, I will use 'understanding a concept', 'possessing a concept', and 'grasping a concept' interchangeably.

[3]Gödel also explicitly claims that some axioms of set theory are analytic in Gödel (1944, 1995). For more on Gödel's notion of analyticity, see for instance Parsons (2014).

[4]This is due, in particular, to famous texts like Quine (1951), Putnam (1962), and more recently Williamson (2006, 2007).

itself—can enable us to recognize that Gödel was right. This is what I will be concerned to show in this paper.[5]

Let me first draw some distinctions to identify and make precise the Gödelian claim I will be defending in this paper. Many contemporary philosophers of set theory have been interested in Gödel's statements for the prospect of arguing for new axioms of set theory.[6] For instance, some philosophers of set theory nowadays try to justify some strong Reflection Principles by saying that these principles "unfold the content of the content of set."[7] My goal, ultimately, is to have something to say about the case for new axioms. But what I think we should do first is step back from those discussions, and carefully examine the simpler case: What does it even mean to say that some axioms "unfold the content of the concept of set," and can we defend this claim for at least the most basic axioms of set theory? This is why, in this paper, I will restrict my attention to what Gödel says about the relation between the concept of set and "the usual axioms" of set theory, namely Zermelo–Fraenkel Set Theory with Choice (henceforth 'ZFC').

In the quotes above, Gödel seems to be saying that understanding the concept of set is sufficient for being in a position to know that ZFC is "implied by" the concept of set.[8] I take this to be equivalent to saying that ZFC "unfolds the content of the concept of

[5]Of note is that I am not aiming for the most historically accurate reading of Gödel's statements. For careful historical treatment of Gödel's statements, see for instance Martin (2005) or Parsons (2014).

[6]See for instance Feferman (1996), Koellner (2009), (Maddy, 2011).

[7]See especially Welch (ms).

[8]Gödel also elsewhere says that ZFC is "intrinsically justified" or is "part of" the concept of set (see for instance Gödel (1944, 1964, 1995)), so it is clear that he intends ZFC to be included among the axioms that "unfold the content of the concept of set."

set," and will call it 'the Gödelian unfolding claim'. As I understand it, ZFC is "implied by" the concept of set in that ZFC at least partly determines the content of the concept of set. Gödel would thus seem to be saying that understanding the concept of set is sufficient for being in a position to know that ZFC at least partly determines the content of the concept of set. It is generally assumed that knowing the content of a concept doesn't imply knowing that there are things that fall under that concept.[9] Hence, on a weaker (and perhaps more plausible) reading, Gödel would seem to be saying that understanding the concept of set is sufficient for being in a position to know that if there are sets, then they satisfy ZFC. Let the *conditional weakening* of an axiom, A, be of the form 'If sets exist, then A'.[10] The Gödelian unfolding claim implies:

Unfolding Understanding the concept of set is sufficient for being in a position to know the conditional weakenings of (at least some of) the axioms of ZFC.[11]

For now, I will use the expression 'in a position to know' liberally, such that one is in a position to know something just in case, if one doesn't already know it, then one could come to know it after carefully reflecting on one's concept. I will later make it more

[9]See for instance Boghossian (2003a), Jackson (1998a), or Williamson (2007). There is some evidence that Gödel himself thought this, see for instance Martin (2005) for discussion.

[10] I will leave it open how exactly to construe the conditional weakening of an axiom. It may also be construed as, for instance, 'If there is a property, S, and a relation, $\in$, that satisfy ZFC, then A', etc. What will matter is that, however one construes the conditional weakening of A, believing the antecedent of the conditional weakening of A is tantamount to assuming that sets exist for the purpose of solving a basic mathematical problem (see Section 1.1).

[11]I intend Unfolding to be necessary. One can argue for a constitutive (and not merely modal) link between concept possession and knowledge on the basis of Unfolding, but this is a task I set aside in this paper.

precise what I mean by 'carefully reflecting on one's concept' (in Section 1.2).

We may now distinguish between two versions of Unfolding, one more modest than the other. The modest version of Unfolding is that possessing the concept of set today, i.e. after set theory has been axiomatized, is sufficient for being in a position to know the conditional weakenings of (at least some of) the axioms of ZFC. The ambitious version of Unfolding is that possessing the concept of set that was in place before set theory was axiomatized is sufficient for being in a position to know the conditional weakenings of (at least some of) the axioms of ZFC.[12]

In this paper, I will leave it open whether there is one concept of set used before and after the axiomatization of set theory, and I will concern myself with the modest version of Unfolding, although some of my discussion will also bear on the plausibility of the ambitious one (see Section 1.2.2). The claim I am interested in is modest in that it doesn't help us give an account of how we can (could have) come to discover the axioms of set theory in the first place. It is nonetheless ambitious in that it provides a necessary condition on possessing the concept of set; it is thus part of a general account of what it is to possess the concept of set, and may help us give an account of what it is to possess mathematical concepts more generally. Moreover, and as I mentioned earlier, the claim I am interested is immodest in that it implies the existence

[12]Given that Gödel was interested in discovering new axioms, he most likely wanted to make the ambitious claim. But if so, he should also be committed to the modest claim, given that he seems to think that the concept of set that was in place before set theory was axiomatized is the same as the concept of set today (see for instance Gödel (1995, 302)).

of conceptual truths, understood here as propositions (whose conditional weakenings) one is necessarily in a position to know if one possesses the relevant concept. I will not explain the exact relationship between Unfolding and analyticity in this paper (though see Sections 1.1.3 and 1.3), but my arguments for Unfolding will lay the foundation for an account of analyticity in mathematics, which is one of the background motivations of the paper.

Here is the plan. I will first argue for Unfolding on the basis of our ordinary practice of ascribing mathematical concepts. According to that practice, a speaker is credited with having the concept of set only if she is able to use some of the axioms of ZFC in the course of solving mathematical problems. We will see in detail in Section 1.1 how that observation leads to Unfolding. This is an important argument to make, because our theorizing about concepts should be responsive to our practices using these concepts. Nonetheless, what exactly it takes to possess a concept has been the topic of a great deal of research, with some quite revisionary positions being defended in the literature. Against this backdrop, simply relying on our ordinary concept-ascribing practice may appear too simplistic. So, in Section 1.2, I will approach the task of defending Unfolding again, this time from a theoretical point of view. I will examine three main accounts of concept possession, some of which might seem to go against the ordinary practice of ascribing mathematical concepts, and show that Unfolding is true on these accounts, at least if these accounts are to be in line with the way actual people possess the concept of set and with the actual practice of set theory. I will conclude that both the ordinary practice of ascribing mathematical concepts, and the mathematical practice itself, come together in revealing that cognition of sets—unlike, perhaps, cognition of non-mathematical objects—is necessarily linked to having access to a fixed set of claims about sets, namely some axioms of set theory.

1.1 From the practice of ascribing concepts to Unfolding

Let me first make explicit the assumptions that will be in the background of my discussions throughout the paper, and introduce some terminology. I will assume that there is a unique concept of set shared by people doing set theory today, i.e. after set theory was axiomatized in the early twentieth century. Call it 'SET'. SET is the meaning of the expression 'set' used by these people.[13] I will assume, as it is commonly done concerning concepts generally, that SET and 'set' have (or determine) a unique content. I won't take a stand here on how exactly to characterize this content. One may take it to be an intension, i.e. a function from possible worlds to extensions. Given that sets—like mathematical objects more generally—exist necessarily if they do, one may also simply take the content of SET to be an extension (the collection of all sets in the actual world). In any case, what I will assume is that the axioms of ZFC at least partly determine what is part of the content of SET. So, for instance, if the content of SET is an intension, then this means that, at every possible world, the value of the intension of SET at that

[13] Throughout this paper, I will use 's means C by *e*', 's's expression, *e*, expresses C', and 's has C' interchangeably. For simplicity, I will ignore concepts that aren't expressed by linguistic expressions.

world contains at least the sets in the cumulative hierarchy of sets, V, that are said to exist by ZFC (an empty set, the set of the empty set, the powerset of the empty set, etc). So ZFC is consistent, and there is a property, S, and a relation, $\in$, which satisfy the ZFC axioms. Call this set of assumptions, 'the Default Position'.

Just to be clear, notice that it is not part of the Default Position that possessing SET puts one in a position to know the content of SET. For all the Default Position says, SET might be like the "folk" concept WATER: The value of the intension of WATER at every world is the set of all sufficiently large aggregates of H_2O molecules (let us assume), but no amount of reflection on one's concept will by itself reveal this. Notice also that the Default Position says nothing about *how* the content of either 'set' or SET is determined; the Default Position only states what is (part of) their content. Finally, notice that I leave it open whether some parts of the content of SET are indeterminate. So, for instance, for all I have said so far, it is possible that the content of the concept SET doesn't settle whether or not one of the sets is a bijection between the powerset of the natural numbers and the first uncountable ordinal.

The Default Position states the default, standard, straightforward view concerning the concept of set: There are sets, they at least satisfy the axioms of ZFC, and we can have thoughts and talk about them. This is what is assumed in the practice of doing set theory, and Gödel himself would most likely have accepted this assumption.[14] So,

[14] Gödel seems to think that there is an "objective realm" of sets to which we have access by grasping the concept of set. Sometimes, he confusingly states that set theory is about concepts, and not about objects (see for instance Gödel (1944)). It is hard to interpret Gödel's views on

in outline, this paper can be seen as an investigation of what is required for thinking about and talking about sets on the naïve, default view of what it is we do when we think about and talk about sets.[15] One may, of course, question whether there really is a unique concept of set shared by all people who say they do set theory—but there is at least a substantial subgroup of them who are not talking past each other when they do set theory, and this just means that they (at least the people in this sub-community) share a concept. One may perhaps also question whether the content of the concept SET is given by ZFC; perhaps there is some question about some axioms such as Choice, as some recent discussions in set theory suggest (for instance recently in Bagaria et al. (ms)). But for my arguments today, I only want to assume that at least the most basic axioms of ZFC (Extensionality, Pairing, Union, Separation, Foundation) determine the content of the concept SET; it need not be given exactly by ZFC.

My first task here in Section 1.1 is to work to defend Unfolding based on our ordinary practice of ascribing mathematical concepts. For this, I propose the following argument, the conclusion of which implies Unfolding:

1. Possessing SET requires having the ability to solve basic mathematical problems about sets.
2. Having the ability to solve basic mathematical problems about sets requires know-

concepts, but he seems to agree with the claim that the concept of set (as well as, perhaps, the expression, 'set') "latches onto" sets as described at least in part by ZFC, and that by grasping the concept of set, we can come to know that ZFC is true of sets. For more on Gödel's views on concepts and realism, see for instance Martin (2005).

[15]These remarks on the Default Position are inspired by Peter Koellner's position in, for instance, (2006; msb). On the other hand, Hamkins (2012), for instance, can be understood to argue for an alternative to the Default Position on the basis of independence results.

ing the conditional weakening of at least some axioms of ZFC.

3. So, possessing SET requires knowing the conditional weakening of at least some axioms of ZFC.

Let me first make some clarifications. By 'basic mathematical problem about sets', I mean the kind of problem one might encounter at the very beginning of a textbook on set theory. For instance, a basic mathematical problem about sets may ask one to build new sets from old sets (such as a union set, power set, or pair set), or to show that there is a unique empty set, or that every set has a unique power set, or that there is no set of all sets, etc.

Solving some basic mathematical problems about sets requires proving some mathematical statement(s). Throughout, I will take proofs to be informal (as opposed to formal derivations). Moreover, I will assume that when one works to solve a mathematical problem about sets, one implicitly assumes that there are sets—understood here (and throughout) as objects referred to by SET. This assumption plausibly captures the practice of doing mathematics. See for instance the following passage from an introductory textbook on set theory:

> [W]e shall think of the collection of all sets as being a clearly defined notion, and whenever we want to show that a sentence, σ, say, of …[the language of set theory] has a formal proof (from ZF say) we simply give an informal argument that the proposition asserted by σ about this collection is true. (Zilber, ms, 5)[16]

[16]See also for instance Kunen (1980, 6ff.).

Finally, I will take it that assuming that there are sets in this way is tantamount to assuming the antecedent of the conditional weakening of an axiom.

Let me now defend the premises of this argument in turn.

1.1.1 Premise 1

Almost everybody accepts that concept possession goes along with possessing certain cognitive abilities. For instance, psychologists often simply define concepts as the cognitive mechanisms by which we categorize, hence for them possessing a concept implies (at least) having the ability to categorize.[17] Even externalists, who generally impose weak demands on concept possession, agree that concept possession requires having some cognitive abilities. Take for instance Tyler Burge. According to Burge, an individual may possess the concept ARTHRITIS even if they are radically mistaken about what the concept can apply to; even if they think, for instance, that ARTHRITIS also applies to ailments of the thighs. Nonetheless, Burge still thinks that possessing ARTHRITIS requires having *some* associated abilities:

> Having …[the concept of arthritis] requires having certain associated discriminatory abilities. …Thus the individual must be able to discriminate arthritis from such things as animals, trees, and numbers, and from certain other diseases, in order to have the concept. But he need not be able to discriminate it from all other rheumatoidal diseases, actual or possible—except insofar as he does so by employing the concept *arthritis*. (Burge, 1993, 325)[18]

[17]See for instance Prinz (2002, 11) and Machery (2009, 12).
[18]See also Burge (1986).

So it is uncontroversial that possessing a concept goes along with possessing certain cognitive abilities. The more controversial question is: For a given concept, which cognitive abilities are required for possessing that concept? There is an uncontroversial part to the answer to that question: Possessing a concept requires at least having the most basic abilities connected with that concept. And, whatever the general story is about what are the basic abilities connected with a given concept, it is very plausible that the basic abilities connected with *mathematical* concepts are basic *mathematical* abilities. At least in practice, we ascribe mathematical concepts to people only if they display some basic mathematical abilities, where a paradigm mathematical ability is the ability to solve mathematical problems, for instance, the ability to prove some result or to calculate some value. So, for instance, if someone was unable to solve basic addition problems, such as calculating $3 + 9$, we would be reluctant to ascribe to them the concept PLUS. This practice can also be found in the psychological literature on mathematical concepts: Psychologists who study the nature of concept possession and concept learning in mathematics only ascribe concept possession based on the subject's abilities to solve mathematical problems.[19]

And this practice also applies to SET: At least according to our ordinary practice of ascribing SET, possessing SET requires having the ability to solve basic mathematical problems about sets, such as building new sets from old sets (pairs, unions, powers sets,

[19] See for instance Tall (1991) for a collection of psychological studies of mathematical concept possession.

etc.) and proving some basic results about sets (that the empty set is unique, that there is no set of all sets, etc.).[20]

One can argue for Premise 1 even without going through the claim that concept possession in general requires having certain associated abilities. As I defined it above, SET is the concept that set theorists have and that rigidly picks out sets in V said to exist by ZFC. It should seem very plausible that in order to have that very concept, one needs to be able to do some basic set theory. Of course, there are people who haven't taken any course in set theory—and who wouldn't be able to solve basic mathematical problems about sets—who can nonetheless reliably apply 'set' to ordinary collections of objects, who perhaps know some things about Venn Diagrams, etc. But why think that such people thereby have SET? At least set theorists would never think that this is enough to possess SET; in practice, they would never ascribe SET to, say, their student, if the student doesn't have the ability to solve the most basic mathematical problems about sets. One could instead think that such people have a different concept, say SET_O ('O' for 'ordinary'), which they express by 'set'. One could think that SET_O applies to ordinary collections of objects, and that either SET_O doesn't apply to all sets in V

[20] It may be that possessing SET doesn't require having the ability to solve all basic mathematical problems about sets, but only sufficiently many, or only some basic mathematical problem about sets or other. I think that at least in practice, possessing SET requires having the ability to solve all *basic* mathematical problems about sets, and I allow that there may be very few such basic problems. In any case, weakening Premise 1 in this way would not make a difference to my arguments here in Section 1.1, as for the overwhelming majority of basic mathematical problem about sets, there will be some axiom (often Extensionality) whose conditional weakening needs to be known in order to have the ability to solve those problems, by an argument similar to the one I give at the end of Section 1.1.2.

guaranteed to exist by ZFC (such as $\aleph_\omega$), or that it is indeterminate whether or not SET$_{O}$ applies to all such sets in V. If one wants to deny Premise 1 and argue that people who don't have the ability to solve basic mathematical problems have precisely SET, and not some other concept like SET$_{O}$, then one needs to give some theoretical reasons for this: One needs to argue that such people can have SET in virtue of standing in some privileged relationship to sets, set theory, or set theorists. I will examine such theoretical reasons in Section 1.2.

1.1.2 Premise 2

I will defend Premise 2 by arguing that having the ability to solve basic mathematical problems about sets requires knowing the conditional weakening of Extensionality. This will provide the template for an argument that may apply to other axioms of ZFC. I will briefly explain why this template should also apply to other axioms of ZFC. As I mentioned in the introduction, my background motivation is to defend the existence of conceptual truths; thus even showing that Unfolding holds for a small part of ZFC would be an achievement in this respect.

Extensionality

Consider the following problem:

Unique Show that there is a unique empty set.

Unique is a basic mathematical problem about sets. It is a simple problem that occurs at the very beginning of an introductory textbook on set theory (cf. Koellner (2010, 6)). In practice, anyone who counts as able to do set theory would be expected to be able to solve this problem.

I now want to argue that one cannot have the ability to solve Unique if one doesn't know the conditional weakening of Extensionality, i.e. that if there are sets, then if a and b are sets, then $a = b$ just in case a and b have exactly the same members.

First, note that one can prove that one cannot prove that there is a unique empty set without assuming Extensionality.[21] Assume now that Alice is trying to solve Unique, but that she doesn't know the conditional weakening of Extensionality. From my assumption above, Alice first implicitly assumes that there are sets. She then starts her reasoning as follows: "I need to show that there is no more than one empty set. So let me assume that sets e and e' are empty. I should show that $e = e'$." I assume that to know that Extensionality is true on the assumption that there are sets just is to know the conditional weakening of Extensionality.[22] Alice assumed that there are sets. There now two cases to consider:

Case 1 Alice doesn't believe that if a and b are sets, then $a = b$ just in case a and b have exactly the same members;

Case 2 Alice believes but doesn't know that if a and b are sets, then $a = b$ just in case

[21] This is because one can construct a model of ZFC $\setminus$ Extensionality with, say, $\{x \mid x \neq x\} \neq \{x \mid x \text{ is a unicorn}\}$.

[22] If one has doubts about this, then one can re-construe conditional weakenings such that this is true. Cf. footnote 10.

a and b have exactly the same members.

Take Case 1 first. If Alice doesn't believe that sets are identical just in case they have the same members, then she won't be able to show that $e = e'$. In order to show that $e = e'$, she must be able to conclude that $e = e'$ from the assumption that e and e' have the same members. If she doesn't believe that, in general, sets a and b are identical just in case they have the same members (either because she has no beliefs about this or because she believes there are exceptions to this), then she won't be able to prove that e must be identical with e'.[23]

Alice also doesn't have the cognitive ability to solve Unique. This is because her inability to show that there is a unique empty set in this case isn't due to some temporary external impediment or psychological shortcoming (such as an aversion to empty sets). So it cannot be argued that Alice has the ability to prove that there is a unique empty set even though she is unable to prove it in this instance.

[23]Of course, strictly speaking, Alice could "solve" Unique if she thinks, say, that only sets with no or only finitely many members obey Extensionality. In that case, she could argue from the assumption that e and e' have the same members (because they have no members), to the claim that $e = e'$ by using such a restricted version of Extensionality. There are two ways to avoid this kind of counterexample. First, we can simply choose a different problem, such as the problem of showing that Replacement yields a unique set; that is, for a given definable function, F, and a given set a, the problem of showing that there is a unique set b that is the image of that a under F. This is also a very basic mathematical problems about sets, and since it is a general claim about all a and F, one needs full Extensionality in order to prove it. A second way to avoid such counterexamples is to argue (as I will at the end of this Section) that having the ability to solve mathematical problems requires gaining knowledge: It requires *coming to know* the solution to the problem. Now if Alice "solves" Unique by using a weaker principle of Extensionality (one that only applies to sets with no or only finitely many members), then her case would be a Gettier case: Alice wouldn't count as *coming to know* the solution to Unique, she would just luckily stumble upon it based on wrong principles.

Now consider a different scenario that may be thought to exemplify Case 1.[24] Assume that Berna's odd philosophical views commit her to believing that there are sets, but that they aren't extensional. She thinks that ordinary set theorists who believe that sets are extensional are wrong about the "true nature" of sets. Assume, moreover, that Berna is nonetheless able to solve these basic mathematical problems by reasoning as follows: "Let me go along with ordinary set theorists and assume for the sake of argument that there are sets, that these are extensional, etc. Of course, ordinary set theorists are wrong about *sets*. So it is better to call what they purport to talk about 'shmets', as these are not really sets. Now I can show that there is a unique empty shmet." This scenario might seem to provide a counterexample to Premise 2: Berna seems able to solve Unique without believing that if there are sets, they are extensional. However, notice that Berna assents to 'if there are shmets, then shmets are extensional'. Moreover, I have assumed that Berna uses 'shmet' to talk about the sets that set theorists purport to talk about, hence, assuming she can do so successfully, the content of 'shmet' is the same as the content of 'set' used by set theorists. Now by the Default Position, the content of 'set' used by set theorists just is the content of SET. So, little does Berna know, the content of her expression 'shmet' is the same as the content of SET. I will discuss the case of possession of SET by deference in more detail in Section 1.2.3. But for now, we can say that on the most natural way of understanding this scenario, Berna

[24] Williamson (2006, 2007) proposes these kinds of cases as counterexamples to understanding–assent links.

isn't a counterexample to Premise 2: She does believe that if there are sets, then they are extensional. She simply expresses SET by 'shmet' instead of 'set'.

Now consider Case 2. It is very plausible that the solving a basic mathematical problem requires gaining some knowledge. In general, it is very plausible that the abilities associated with concept possession are abilities to gain knowledge (so for instance, the ability to discriminate or categorize things as arthritis is an ability to gain knowledge about whether some particular object, like the number 7, is arthritis). Since, as I just argued, having the ability to solve Unique requires believing Extensionality, if the reasoning that Alice undergoes is to give her knowledge of the solution to Unique, then she must also know the premises of her argument for solving Unique. So Alice must not only believe but know the conditional weakening of Extensionality in order to have the cognitive ability to solve Unique.

I just argued that having the ability to solve Unique requires knowing the conditional weakening of Extensionality. Since Unique is a basic mathematical problem about sets, the argument I provided supports Premise 2.

Other axioms

The template of the argument in Section 1.1.2 will apply to other axioms of ZFC depending on which are the basic mathematical problems about sets. In order to find out what are basic mathematical problems about sets, we need to look at the practice of set theorists, i.e. we need to look at what is in textbooks, what teachers usually teach to beginner set theorists, etc. For instance, problems that ask one to build new sets out of old sets (such as power sets, unions, intersections, singletons, etc.) are basic mathematical problems about sets. So knowing the conditional weakenings of at least Pairing, Union, Power Set, Empty Set, and perhaps Separation will be a necessary condition for possessing set by the argument for Unfolding.[25]

1.1.3 Preliminary conclusions

In Sections 1.1.1–1.1.2, I argued that, if we take into account the intuitive, ordinary practice of ascribing mathematical concepts, possessing SET requires knowing the conditional weakening of at least some axioms of ZFC. This, in turn, implies Unfolding.

This argument showed something specific to formal mathematical concepts. Take, for instance, the concept ARTHRITIS. If we agree with Burge above, then, according to the ordinary practice of ascribing ARTHRITIS, the set of abilities required for possessing ARTHRITIS contains (let us say only) the ability to discriminate arthritis from trees, animals, numbers, and chickenpox. One can have these abilities either by knowing that arthritis is an inflammation of the joints (and that trees, numbers, animals and chickenpox aren't), or by knowing that arthritis generally causes joint pain (and that trees, numbers, animals and chickenpox don't generally cause joint pain), or by knowing

[25]There might be some vagueness concerning what counts as a basic mathematical problem about sets. If so, then there will also be vagueness concerning what is required for possessing SET, and hence vagueness concerning whether one possesses SET (as opposed to expressing a different concept by 'set'). This kind of vagueness is hard to avoid for any account of concepts, and it is compatible with the assumption in the Default Position that there is a unique concept, SET, that is the meaning of 'set' used by the community of set theorists.

that one has arthritis in one's finger (and that one doesn't have trees, animals and chickenpox in one's finger), etc. So for a concept like ARTHRITIS, there is no fixed set of information about arthritis which one needs to possess in order to have the abilities required for possessing ARTHRITIS. The conclusion of the argument above is that ARTHRITIS and SET are different in precisely this respect: There is a fixed set of information which one needs to possess in order to have the abilities required for possessing SET; this set contains (at least) some axioms of ZFC. Both the set of abilities required for possessing ARTHRITIS, and the set of abilities required for possessing SET, contain the most basic abilities required to possess ARTHRITIS and SET respectively; for ARTHRITIS, these are basic discriminatory abilities, and for SET these are basic mathematical abilities. But only possessing the abilities required for possessing SET requires knowing a fixed set of propositions.

The argument above thus brings out a special feature of mathematics and mathematical concepts. Unlike the domain of diseases or physical objects—and, perhaps, more generally, unlike the domain of empirically discernible entities—there is one "canonical" way of having cognition of the domain of formal mathematical objects. That is, for a formal mathematical concept, C, one must know a certain fixed set of principles in order to have cognition of the domain of things that fall under C. The domain of empirically discernible entities, on the other hand, is "multiply accessible."

I reached these conclusions based on our ordinary practice of ascribing mathematical concepts. But there are some accounts of concept possession that go against our practice of ascribing mathematical concepts; that is, there are some theoretical reasons to reject Premise 1. In Section 1.2, I will examine these accounts and argue for Unfolding again, this time from a theoretical point of view.

Before I do so, let me say a few words about the relationship between the argument I presented here in Section 1.1, Unfolding, and conceptual or analytic truths.

One might wonder whether a conceptual truth, relative to a given concept, should be defined as a proposition one is necessarily required to know in order to possess that concept, or whether it could be defined (more weakly) as a proposition one is merely in a position to know in order to possess that concept.[26] This, of course, will depend on what it is exactly to "be in a position to know" something, which I will explain in more detail in Section 1.2. I think that both of these are valid definitions of 'conceptual truth', for reasons I will become clearer in Section 1.2. But notice that the argument template I proposed here in Section 1.1 shows that there are conceptual truths in the strong sense. The argument template is so strong that it can be used to show that knowing some propositions besides the conditional weakenings of the axioms of ZFC is also a requirement on possessing SET, for instance some logical laws. But this is not a great concern for the present argument, since one can either define 'analytic' or 'conceptual' truths as a proper subset of the propositions (whose conditional weakenings) one is required to know for concept possession (for instance the subset that is in some

[26]On some understandings of what it is to be 'in a position to know', Timothy Williamson, for instance, would most likely only accept the former definition. See for instance Williamson (2007, 130ff.).

sense "essentially" about sets), or grant that one needs logical knowledge (and perhaps possession of logical concepts) in order to possess SET.

There is a lot more to be said about the exact relation between analyticity, the argument above, and Unfolding, and I set this task aside in this paper. But the argument here in Section 1.1 at least provides some of the foundations for an account of analyticity or conceptual truth in mathematics.

1.2 From theories of concept possession to Unfolding

From a broad perspective, there are four general philosophical theses concerning what determines that one possesses a concept: Descriptivism, Causal Externalism, Reference Magnetism, and Social Externalism. These can either be seen as mutually incompatible accounts of what it is to possess a concept simpliciter, or they can be seen as mutually compatible accounts of different *ways* of possessing a concept. This won't make a difference to my arguments, although for simplicity I will often adopt the latter view.[27]

Causal Externalism very plausibly won't apply to mathematical concepts such as SET. Given this, here are the main available accounts of what determines that one possesses SET: One can possess SET by associated description, by deference to an optimal theory, or by deference to other people. Deference to an optimal theory is a version of Reference Magnetism: Reference Magnetism isn't really by itself a metasemantic theory, it is

[27]Mark Greenberg (2014a; 2014b) recently argued that one must supplement such a view (according to which there are different ways of possessing a concept) with a general account of what content should be such that it could be determined by these various factors.

instead the view that, given certain other constraints, those things that are "magnetic" get picked out by a particular mental state.[28] The most plausible way of making sense of this general idea in the case of a mathematical concept is to take the "magnets" to be properties described by an optimal theory, given certain constraints on that theory. I will explain this view in more detail shortly. Call the claim that these are the only three available accounts of possession, 'Exhaustiveness'.

As will become clear, both the second and third accounts can provide theoretical motivation to reject Premise 1. In what follows, I will argue that each of these accounts nonetheless validates Unfolding. In broad outline, my strategy in each subsection will be as follows. I will first look at the account of concept possession in the abstract, and see how close one can get to Unfolding from there. As we will see, Unfolding follows from the first account immediately, but not quite so immediately from the second and third accounts. I will then show, based on some observations about the actual practice of set theory, that Unfolding must be true on these accounts, if these accounts are to explain how actual people possess SET, and if they are to fit with the existing practice of set theory.

So my aim here in Section 1.2 isn't to show that it is metaphysically impossible to possess SET without being in a position to know the conditional weakenings of at least

[28]Reference Magnetism is a view often attributed to Lewis (1983). In outline, according to Reference Magnetism, the fact that speaker s means SET by 'set' is determined at least in part by the "naturalness" of the content of SET. More more about the relationship between Reference Magnetism and Deference to an Optimal Theory, see fn. 33.

some of the axioms of ZFC. Rather, what I want to show is that, according to the three main accounts of possessing SET—and especially when these provide accounts of the way actual people possess SET and fit with the existing practice of set theory—possessing SET in the way proposed by these accounts requires being in a position to know the conditional weakenings of at least some of the axioms of ZFC.

1.2.1 Associated Description

Consider, first, the following account of possession of SET:

Associated Description One can possess SET by associating some description, D, with 'set'.

Associated Description is a "descriptivist" account of concept possession, along the lines of traditional descriptivist accounts of reference, and more recently defended by, for instance, Frank Jackson (1998a; 1998b) and Christopher Peacocke (1998; 2008). Following Jackson, I will assume that a *description* is a property, where a property is understood in the liberal sense of being a way some thing might be Jackson (1998b, 202). And I will assume that a description one *associates* with an expression is something that guides one's application of the expression. As such, one's associated description need not be available "before one's mind," and one need not be able to explicitly articulate one's associated description, but one can in principle come to know one's associated description if one reflects on how one applies the expression.

Following an analogy given by Jackson, the relationship one has to one's associated description is, I will assume, like the relationship Charlotte has to the recursive definition of well-formed formulas: Charlotte can say, for any given formula, whether or not it is well-formed, and why (for instance, presented with '$(p \supset q$', she can say that this is not a well-formed formula because the right parenthesis is missing), but she cannot give an account in words that covers all the cases Jackson (1998b, 211f.). If Charlotte had enough time, and if she reflected on what guides her judgments in each case, she could, in principle, list all and only the well-formed formulas. In that sense, the list of well-formed formulas is "in her mind." Charlotte cannot state the pattern that underlies this list, which, in this case, is the recursive definition. But the recursive definition is a priori accessible to her, that is, with some ingenuity, she could in principle come to extract the pattern that underlies her judgments in each case, even though this may take a lot of effort. As I will understand the expression, Charlotte is in this way "in a position to know" the recursive definition of well-formed formulas. I will assume that one is, in the same sense, "in a position to know" one's associated description.[29]

I am thus working with a weak version of descriptivism: one that doesn't require the thinker's associated description to be "before the thinker's mind" or even easily accessible to the thinker. In this way Associated Description captures the common core

[29]This view of associated descriptions is also in line with Peacocke's understanding of "implicit conceptions" as states of "tacit knowledge" required for concept possession and that guide one's deployment of the concept (see for instance Peacocke (2008, 113ff.)). Of note is that, unlike Charlotte, one need not have a judgment about how to apply one's expression in every possible case: One's associated description is what underlies one's application of the expression, whatever that application is, even if it is indeterminate in many cases.

of many descriptivist accounts of concept possession.[30]

If a thinker, s, possesses precisely SET by associating some description, D, with 'set' (henceforth simply 'by associated description'), then D must determine what is the content of the concept SET. This is because s's associated description, D, determines that s possesses SET, and SET uniquely determines a content. By my assumptions above, s is in a position to know D. Is s also thereby in a position to know the content of SET?

Not necessarily. To see why, consider the following case. Some philosophers have argued that we can possess WATER by associating a description with 'water' roughly like 'The actual watery stuff here'.[31] These philosophers argue that this associated description determines the content of WATER, in that, according to the ordinary semantics of 'actually', 'The actual watery stuff here' refers to sufficiently large aggregates of H_2O molecules at every possible world. Even though this associated description determines the content (let's say in this case the intension) of WATER, there is no way for a thinker who associates this description with 'water' to know this merely by reflecting on their pattern of application of 'water': The thinker needs additional empirical information to determine what is the chemical composition of the actual watery stuff here.

There are a number of ways in which a description can in this sense "opaquely" determine the content of the concept SET—opaquely in that the thinker isn't in a position to

[30] Associated Description thereby encompasses many specific versions of descriptivism. In particular, Associated Description can encompass various versions of epistemic two-dimensionalism (cf. for instance Chalmers (2004)), or George Bealer's account of concept possession (cf. for instance Bealer (1998)).

[31] See for instance Lewis (1995) or Jackson (1998a).

know the content of the concept, as it is given by ZFC (or something logically equivalent to ZFC), just because she has that associated description. For instance, one might possess SET by associating the description: 'Whatever set theorists talk about when they do their mathematical work'. Here, I will treat cases of possession of SET by opaque associated descriptions separately in the next two sections, as cases of concept possession by deference to an optimal theory and by deference to others. I do this because even though the cases are different from a metasemantic perspective, they are epistemically equivalent.

So, if we restrict Associated Description only to transparent (i.e. non-opaque) associated descriptions, then this first account of concept possession implies Unfolding for this particular way of possessing SET: Necessarily, whoever possesses SET by transparent associated description is in a position to know the content of the concept of SET, that is, by the Default Position, whoever possesses SET by transparent associated description is in a position to know the conditional weakenings of ZFC.

I just argued for Unfolding on the basis of Associated Description, without considering how Associated Description would provide an account of how actual people possess SET. If we do take this into consideration, however, then we can conclude an even stronger version of Unfolding. Indeed, ordinary set theorists who possess SET actually have many of the axioms of ZFC "before their mind": They know what are at least the most basic axioms of ZFC. So, if they possess SET by associating a transparent description with 'set', then at least a large part of the description they associated with 'set' must be "before their mind." Thus if set theorists possess SET by transparent associated description, then, in order to possess SET *in the very same way* as they possess it, one must know (and not merely be in a position to know) at least the majority of the axioms of ZFC. We see here that when we apply the first account of concept possession to the way actual set theorists possess SET, we get an Unfolding claim concerning that particular way of possessing SET that is as strong as the Unfolding claim I argued for in Section 1.1 on the basis of our actual practice of ascribing SET.

1.2.2 Deference to an Optimal Theory

One externalist thesis is that for some concept, C, some individual, s, and some set of constraints or desiderata, Γ, the fact that s possesses C depends on the content of the theory that best satisfies Γ. In such cases, let us say that s possesses C *by deference to the theory that best satisfies* Γ. Here I intend 'constraint' to be understood in a liberal sense. So, for instance, I will take it that it is a constraint on an optimal theory of derivatives that the theory be a mathematical theory. Let us also say that someone possesses C *purely* by deference to an optimal theory if she possesses C by deference to an optimal theory and not by deference to other people or by associated (transparent) description. (Henceforth when I talk of associated descriptions, I always mean transparent associated descriptions.)

For instance, Georges Rey proposes that Newton and Leibniz had the concept LIMIT by deference to an optimal theory:

> Thinkers construct concepts that have to satisfy certain *constraints*, but for which they have no defining conceptions at all, explicit or implicit. ...Newton and Leibniz postulated infinitesimals and limits without knowing their precise nature, or, in their case, even how to rid the postulation of paradox. We may do them the justice of attributing them the same concept, leaving it a matter of theory what the best explication of that concept might be.
>
> ...Thinkers *postulate* that there is some phenomenon that happens to play (most of) the explanatory roles that interest them, whose nature would be provided by some optimal theory of the phenomenon. (Rey, 1998, 98)[32]

These passages suggest that people who possess a concept by deference to a theory that best satisfies Γ must be in a position to know Γ, i.e. if they reflect on what guides their application of the concept, they can come to know that they apply the concept to things that best satisfy Γ. One obstacle for this claim is that one may have a concept by deferring to a theory that best satisfies somebody else's constraints. So, for instance, it seems possible for Leibniz's student to have the concept LIMIT by intending to have the concept that satisfies Leibniz's constraints on an optimal theory of limits, without himself being in a position to know what those constraints are. But here let us set aside these as cases of impure deference to an optimal theory, because they involve deference to other people. With this clarification in place, I will thus make the following assumption in line with Rey:

(I) An individual possesses a concept, C, purely by deference to an optimal theory with respect to Γ only if she is in a position to know Γ.[33]

[32] See also Smith (2015) for a recent discussion of Rey's view applied to Leibniz and Newton's concept DERIVATIVE.

[33] Assumption (I) rules out strong forms of reference magnetism, according to which what is most "natural" determines that one possesses a concept independently of one's intentions to

Now the idea of deference to an optimal theory suggests the following account of possession of SET. There is a (perhaps not unique) set of constraints, Γ_{SET}, such that:

Deference to an Optimal Theory One can possess SET by deference to a theory that best satisfies Γ_{SET}.

For a subject to possess precisely SET by deference to a theory that is optimal with respect to Γ_{SET}, the constraints, Γ_{SET}, have to be satisfiable by a unique theory. According to the Default Position, that theory must at least include ZFC.

At least on the face of it, Deference to an Optimal Theory is incompatible with Premise 1, as it suggests that one can have SET without knowing anything about what is the optimal theory of sets, let alone having the ability to solve basic mathematical problems about sets.

In order to figure out whether Deference to an Optimal Theory threatens Unfolding, we should take a closer look at the constraints Γ_{SET}. On the most natural proposal, these constraints are purely mathematical and conceptual constraints. For instance, in a recent paper Antony Eagle argues that the ZFC axioms are the "simplest, best, most compact, and most systematic description" of the core platitudes concerning sets (Eagle, 2008, 77), where platitudes concerning sets are sentences that are seen as obvious and

use the word in a particular way. Strong versions of magnetism have many counterintuitive consequences, discussed for instance in Schwarz (2014). Here I set those views aside, but I do admit of a weaker form of magnetism: I allow, for instance, that it may be a constraint on an optimal theory of sets that the theory be "natural." Presumably, which mathematical theory is most natural is something that one is in a position to know (a priori), but even if not, the arguments at the end of this Section 1.2.2 aren't affected by admitting such naturalness constraints in Γ_{SET}.

undeniable by any competent user of the concept (Eagle, 2008, 72). More generally, it is highly plausible that one is in a position to know what satisfies a given set of mathematical and conceptual constraints, only perhaps with a lot of effort. So assuming the Default Position, (I), and the natural view that Γ_{SET} only contains mathematical and conceptual constraints, we can conclude that possessing SET by deference to the theory that best satisfies Γ_{SET} requires being in a position to know the conditionally weakened axioms of ZFC, which gets us close to Unfolding.

How close, though? There are two reasons to think that we're not quite there yet. First, it is not entirely clear that we are in a position to know what satisfies an optimal mathematical theory in the same sense as one is in a position to know one's associated description: The former seems to be much more cognitively demanding than the latter. Second, what satisfies some of the constraints in Γ_{SET} might be contingent; there might be constraints such as 'Integrate your theory with the best physical account of x'. If some of Γ_{SET} does contain such contingent constraints, then being in a position to know the conditional weakenings of ZFC need not be a requirement on possessing SET by deference to the theory that best satisfies Γ_{SET}. Can Γ_{SET} contain such "contingent" constraints?

The optimal theory of sets is a mathematical theory, and so the majority of the constraints on this theory will be mathematical. If some part of Γ_{SET} is contingent, then the relevant constraints can only be comparatively weak ones such as 'Integrate your theory with developments x'. It is therefore highly unlikely that the contingent parts of Γ_{SET} by themselves will be able to determine that, say, sets are extensional, or that there is a set that contains any given two sets. Perhaps the contingent parts of Γ_{SET} will determine some more complicated axioms, such as the axiom of Choice. Indeed, it has been argued that our justification for believing the axiom of Choice is a posteriori, because it depends on the utility of Choice for solving certain mathematical problems (cf. Parsons (1977)). But the non-contingent parts of Γ_{SET} will most likely determine at least the most basic axioms, such as Extensionality, Pairing, and Foundation. So we have reason to think that whoever possesses SET by deference to an optimal theory is in a position to know the conditional weakenings of at least the most basic axioms of ZFC.

Of course, these considerations are not conclusive: It is hard to argue that it is simply impossible to possess SET by deference to an optimal theory without being in a position to know at least some conditionally weakened axioms. So now I want to consider how Deference to an Optimal Theory could provide an account of the way people actually possess SET. In other words, I want to ask: What will Γ_{SET} contain, realistically?

Notice that Rey talks about a concept, LIMIT, before it was well understood and formalized in mathematics. It makes sense to defer to an optimal theory if one operates with a notion that is poorly understood. But set theory has been axiomatized, so it is unclear why people, at least the people who know the axioms of ZFC, would still defer to an optimal theory. This would only make sense if they think that ZFC does not represent the optimal theory of sets. There might be people who are inclined to think that there might be an even better axiomatization of set theory, especially with respect to some more controversial axioms like Choice.[34] But at least the vast majority of set theorists would not be inclined to give up the most basic axioms (such as Extensionality, or Pairing, Separation, or Foundation), even if this was the best way to satisfy some other theoretical constraints.

To take a more specific example, consider set theories without Foundation (for instance, New Foundations with Ur-Elements ('NFU')). The vast majority of set theorists don't regard NFU as an alternative (possibly more optimal) theory of sets of the kind that Gödel and Cantor were talking about (assuming there are such sets). Non-well founded set theories are seen as alternative theories, and, at least according to the vast majority of set theorists, the subject matter of these theories is distinct from the subject matter discussed in standard set theory. This case is even clearer when made for Extensionality.

The fact that set theorists would not be willing to give up these axioms, in turn, suggests that they are not deferring to an optimal theory that isn't constrained by at least the most basic axioms of ZFC. This is because it is very plausible to think that if someone possesses a concept by deference, then they must intend to or be disposed to defer. If someone has no disposition to accept what an optimal theory says, or to revise their beliefs according to what an optimal theory says if they were to come to learn it,

[34] Current discussions about whether Choice holds for the whole large cardinal hierarchy would exemplify this point (see for instance Bagaria et al. (ms)).

then there is no reason to ascribe her the concept that applies to things described by the optimal theory, instead of ascribing her an idiosyncratic concept that best fits her own usage of and descriptions she associates with her expression. So, realistically, $\Gamma_{\textsc{set}}$ contains the constraint that the optimal theory of sets ought to contain the most basic axioms of ZFC, and thus in order to possess SET in the way that actual set theorists possess it, one must be in a position to know the most basic axioms of ZFC.

1.2.3 Deference to People

One social externalist thesis is that for some concept, C, some individual, s, and some linguistic expression, *e*, the fact that s possesses C depends on *e*'s usage by people other than s. In such cases, it is said that s possesses C *by deference to those other people*.[35] Let us say that s possesses C *purely* by deference to others if s possesses C by deference to others and not by associated description or deference to an optimal theory.

For instance, in the famous thought experiment in Burge (1979), the fact that a thinker, call him 'Burt', possesses ARTHRITIS depends on the usage of 'arthritis' by doctors in his linguistic community. If doctors in Burt's linguistic community had used 'arthritis' to mean THARTHRITIS instead of ARTHRITIS, where THARTHRITIS is a concept that applies to ailments of the thighs as well as the joints, then Burt would

[35] Social externalism originates in Burge (1979). Earlier, Putnam (1975a) argues that facts about other people's usage of an expression can determine its intension. Although Putnam doesn't explicitly extend his thesis to the determination of concepts, one could extend it by assuming that the intension of an expression is the intension of the concept it expresses (which is in line with the Default Position, but not exactly in line with Putnam's own usage of 'concept' in (1975a).)

have possessed THARTHRITIS instead of ARTHRITIS.

We see here that 'usage' in the social externalist thesis should be understood so that *e*'s usage by people other than s determines what concept those people express with *e*; for the idea is that *e*'s usage by people other than s determines that *e* expresses some concept, and then s possesses that very concept by deference to those people.[36]

In order to have some handle on the mechanisms of deference to people, I will make the following assumption throughout this subsection:

(II) Someone possesses the concept expressed by linguistic expression, *e*, (purely) by deference to other people only if she is disposed to use *e* (purely) in line with those people's usage of *e*.

Assumption (II) is plausible because it must be possible for a person to associate an idiosyncratic concept with a given public linguistic expression, *e*. Thus if someone doesn't have the disposition to use *e* in line with the usage of others, then there is no reason to ascribe her the concept others associate with *e*, instead of ascribing her an idiosyncratic concept that best fits her own usage of and descriptions she associates with *e*.[37]

Social externalism thus seems to provide a third account of possession of SET:

Deference to People One can possess SET by deference to other people.

[36]For this reason, some philosophers talk of 'meaning-determining usage' instead of 'usage', at least if one thinks that only a proper subset of the total "use" of an expression determines its meaning.

[37]Jens Kipper makes this argument in Kipper (2012, 91). Saul Kripke presupposes something like (I) in Kripke (1972, 96). More recently and concerning a similar topic, Sheldon R. Smith assumes something like (I) in arguing against George Rey's claim that Leibniz possessed DERIVATIVE by deference to an optimal theory (Smith, 2015, 19).

Deference to People threatens both Premise 1 and Unfolding, because it suggests that people can possess SET without knowing the conditional weakening of any axiom of set theory, let alone having the ability to solve basic mathematical problems about sets. In what follows, I will argue that, at least if we take into account the actual practice of set theory, if there are people who possess SET by deference to others, then there are people who possess SET by associated description or by deference to an optimal theory. Thus we can, by the arguments in Section 1.2.1 and Section 1.2.2, maintain Unfolding concerning those ways of possessing SET.

Call the community of people in which deference happens, i.e. the linguistic community closed under deference, 'the deference community'. I first want to argue for the following claim:

Experts Needed If some people possess SET by deference to others, then there must be some people in the deference community (call them, 'the experts') who possess SET at least in part by associated description or by deference to an optimal theory.[38]

From now on, it will help to work with a toy model. Assume that some deference community consists of three people: Doug, Engin, and Ferdan. Assume that their expression 'shmet' expresses the concept SHMET. Assume that the content of SHMET and 'shmet' are determined only by three axioms, A_1, A_2 and A_3. If all of Doug, Engin,

[38]I don't consider partial deference to an optimal theory, because if a subject has a concept, C, both by deferring to the theory that best satisfies constraints, Γ, and also by deferring to others or by having an associated description, D, then we can take add constraints 'The cs are whatever people P mean by *e*' and 'The cs satisfies D' to Γ to get Γ', and say that this subject has C by deferring to a theory that best satisfies Γ'.

and Ferdan possess SHMET purely by deferring to others in their deference community, then none of them can have SHMET. Assume, for instance, that Doug possesses SHMET purely by deference to Engin, and that Engin possesses SHMET purely by deference to Doug. Since Doug possesses SHMET purely by deference to Engin, what determines that Doug means SHMET by 'shmet' is that Engin means SHMET by 'shmet'; there is nothing else about Doug's usage or knowledge that helps determine that he means SHMET by 'shmet'. But since Engin also purely defers to Doug, no reference (or intension) is grounded: Doug purely defers to Engin, who purely defers back to him. Adding Ferdan will not change anything to the general structure of this case, since Ferdan will also purely defer (either to Doug, to Engin, or to both), and there is nothing about her usage or knowledge that determines she means SHMET by 'shmet'. By Exhaustiveness, there must therefore be someone in the deference community for whom having SHMET is established at least in part by associated description or by deference to an optimal theory.[39]

Applying this model to SET, we thus see that Experts Needed follows: Either there is someone in the deference community of 'set' who possesses SET by deference to an optimal theory, or there is someone in the deference community of 'set' who possesses SET at least in part by associated description. Notice, however, that Experts Needed itself doesn't imply that there must be actual people who possess SET not purely by

[39]It is generally assumed (though often not explicitly argued) that the "chain of deference" must come to an end (see for instance Burge (1979) or Putnam (1975a)). See also Recanati (2000) for a discussion of this principle in a slightly different context.

deference to others. For instance, take a subject, s, in the nearest possible world in which there is no set theory. Presumably, s can possess SET by deferring to people who use ‘set’ in our world (perhaps by associating the following description with set: ‘The concept expressed by ‘set’ in the nearest possible world’). Here, I will set aside those cases because they go against the actual practice of set theory. There actually are experts: There are expert set theorists who know the axioms of set theory (i.e. at least part of the content of SET), and who very plausibly aren’t disposed to give up at least the most basic axioms of ZFC (by the arguments at the end of Section 1.2.2). These expert set theorists thus don’t possess SET purely by deference to others (by (II)). So, if we want to model the actual practice of set theory, we don’t need to look at cases where there aren’t actual experts in the deference community.

Thus, taking into account our actual practice of set theory, Experts Needed tells us there either are some people who possess SET by deference to an optimal theory, or there are some people who possess SET at least in part by associated description. If there are people who possess SET by deference to an optimal theory, then, by Section 1.2.2, their way of possessing SET requires being in a position to know the conditional weakenings of at least the most basic axioms of ZFC. What if there are some people who possess SET at least in part by associated description, and at most in part by deference to others? In the remainder of this section, I want to investigate whether their way of possessing SET also requires being in a position to know the conditional weakenings of some axioms of ZFC.

Let us return to our toy model, and examine the case of people who possess SHMET at least in part by associated description and at most in part by deference to others. There are two cases to consider:

Case 1 (At least) one of Doug, Engin, and Ferdan possesses SHMET by associating A_1, A_2 and A_3 with 'shmet';

Case 2 Doug possesses SHMET by associating A_1 with 'shmet' and by deferring to Engin and Ferdan, Engin possesses SHMET by associating A_2 with 'shmet' and by deferring to Doug and Ferdan, and Ferdan possesses SHMET by associating A_3 with 'shmet' and by deferring to Doug and Engin.[40]

I contend that we don't need to consider a case where no one in the deference community associates either of A_1, A_2, or A_3 with 'shmet'. So for instance, it cannot be that none of Doug, Engin, and Ferdan possess SHMET at least in part by associating A_1 with 'shmet', and yet they (or at least one of them) still possess(es) SHMET. By Exhaustiveness, and given that we've already set aside deference to optimal theories, only deference to people and associated descriptions guarantees reference to shmets. If nobody associates A_1 with 'shmet', then, plausibly, Doug, Engin, and Ferdan should at most count as having a concept, SHMET*, whose intension is determined only by A_2 and A_3. Thus Case 1 and Case 2 are the only ones we need to consider.

[40]For simplicity and without loss of generality, I don't consider the case where some of Doug, Engin, and Ferdan purely defer to others, because such a case will be structurally similar to either Case 1 or Case 2 when we exclude the people who purely defer.

If possession of SET works like in Case 1, we get that someone possesses SET purely by associated description, and so, by Section 1.2.1, the way they possess SET requires being in a position to know the conditional weakenings of the axioms of ZFC.

If possession of SET works like in Case 2, the situation is more complicated. We still get that possession of SET by partial associated description requires being in a position to know the conditional weakenings *some* axiom(s) of ZFC, but we don't get that there are some axioms of ZFC whose conditional weakenings one needs to be in a position to know in order to possess SET.[41] However, there are a number of reasons to believe that Case 2 is incompatible with the actual practice of set theory.

To see this, return to the toy model. Assume for now that if one doesn't possess SHMET by associating some A_i with 'shmet', then one also doesn't know A_i. In that case, neither Doug, Engin, nor Ferdan can do much with their concept SHMET: They cannot solve mathematical problems about shmets, they cannot derive theorems and apply their theory of shmets to other mathematical theories. They won't even know whether their concept is consistent: It may well be that A_1, A_2 and A_3 are jointly inconsistent, but without knowing what all the axioms are, none of Doug, Engin and Ferdan can actually check this. If possession of SET worked like Case 2, then the ordinary practice of mathematics would not be possible.

Now it may be that Doug, Engin, and Ferdan each know all of A_1, A_2 and A_3, but

[41]If ZFC doesn't determine all of the content of SET, then for this claim to be true, we should replace 'ZFC' with the complete axiomatization of set theory instead.

they still instantiate Case 2 because even though, for instance, Doug knows A_2, he defers to Engin and Ferdan concerning A_2. Once again, what we should say here is that, by the arguments at the end of the last section and (II), it is very plausible that most set theorists wouldn't be willing to give up at least the most basic axioms of ZFC, even if this meant satisfying some theoretical constraint, or if some expert set theorist came to reject such an axiom. Thus once again, we may conclude that Case 2 provides an inaccurate model of the way we possess SET. There is more to say and to investigate about the mechanisms of Case 2, especially since, if Case 2 can provide a possible model of how possession of SET works, then we can, on this basis, work to develop a social view of analyticity and a priority. But for the purposes of this paper, they key point to take away here is that, at least realistically, even if some people possess SET by deference to others, some people possess SET purely by associated description or purely by deference to an optimal theory. So we've reduced the case of deference to others, so to say, to the case of associated description and deference to an optimal theory.

1.3 Beyond Unfolding

Let us summarize. In Section 1.1, I argued that possessing SET requires knowing the conditional weakenings of at least some axioms of ZFC. My argument was based on the ordinary practice of ascribing SET. I argued that according to the ordinary practice of ascribing SET, having SET requires being able to solve some basic mathematical problems about sets. I then argued that solving such problems, in turn, requires knowing the conditional weakenings of at least the most basic axioms of set theory.

In Section 1.2, I approached the task of defending Unfolding this time from a theoretical point of view: I examined three main philosophical accounts of concept possession, and argued that Unfolding is true on these accounts, at least if these accounts are to be in line with the way actual people possess SET and with the actual practice of set theory. More specifically, I argued for the following four claims. First, in Section 1.2.1, I argued that possessing SET by (transparent) associated description requires being in a position to know the conditional weakenings of the axioms of ZFC. Second, in Section 1.2.2, I argued that possessing SET by deference to an optimal theory very plausibly requires being in a position to know the conditional weakenings of at least some basic axioms of ZFC. Third, again in Section 1.2.2, I argued that if actual people possess SET by deference to an optimal theory, then their particular way of possessing SET by deference to an optimal theory requires being in a position to know the conditional weakenings of at least some basic axioms of ZFC. Fourth and finally, in Section 1.2.3, I argued that, if we take into account the actual practice of mathematics, then if there are some people who possess SET by deference to others, then either there are people who possess SET by (transparent) associated description, or there are people who possess SET by deference to an optimal theory. The claims in Section 1.2.1 and Section 1.2.2 then apply to these people.

The conclusion of the argument in Section 1.1 implied Unfolding, and the conclusions of the arguments in Section 1.2 implied Unfolding concerning a particular way of possessing SET (or, equivalently, on a particular account of possessing SET). Of note is that Gödel himself should probably be understood as making a claim about a particular way of possessing SET; more specifically, he should probably be understood as making a claim about experts' way of possessing SET. Gödel, after all, was interested in how understanding of the concept of set has consequences when doing set theory. He most likely wasn't interested in whether it is metaphysically possible to possess SET without being in a position to know the conditional weakenings of the axioms. In the quotes at the beginning of this paper, Gödel also invokes a distinction between different "depths" of possessing the concept of set. One could thus think that by different "depths" of possessing the concept of set, he means (or should mean) different "ways" of possessing the concept. Unfolding should, then, be understood as the claim that "deep" possessing of SET requires being in a position to know the conditional weakenings of at least some of the axioms of ZFC.

I leave the task of specifying the relationship between analyticity, Unfolding, and the arguments in Section 1.2 for future research. Of note is that just as one should think, as I mentioned just above, that Unfolding is in some sense restricted to a particular way of possessing SET, one should probably also think of analyticity or conceptual truth as somehow restricted to a particular way of possessing concepts. After all, if there really are different ways of possessing concepts, then it should be no surprise that the notion of analyticity or conceptual truth should also be in some sense "relative" to ways of possessing concepts. So, for instance, one can say that a proposition is conceptually true on a particular way of possessing the concept only if knowing (the conditional weakening of) that proposition is required on that way of possessing the concept. This relativized notion of analyticity or conceptual truth will be one key component of the defense of analyticity or conceptual truth on the basis of the arguments for Unfolding in this paper.

2

Defending Formal Analyticity

Contemporary discussions of analyticity have largely focused on sentences of natural language, such as 'Bachelors are unmarried' and 'Vixens are female foxes', and on basic rules for inferences carried out in natural language, such as modus ponens and conjunction elimination.[1] This is a surprising development from the perspective of the philosophical tradition this debate addresses: Starting at least with Gottlob Frege (1884), traditional defenders of analyticity primarily focused on abstract sciences like mathematics and formal logic, and took axioms and sentences stating rules of inference for formal mathematical languages to be some of the most plausible candidates for analytic sentences.[2] Logical empiricists, in particular, introduced their notion of analyticity in order to account for the necessity and apriority of mathematics from an empiricist standpoint.[3] Rudolf Carnap—one of the staunchest defenders of analyticity—went so far as to argue that analyticity applies only to sentences of formal languages:

> [T]he analytic–synthetic distinction can be drawn always and only with respect to a language system, i.e., a language organized according to explicitly formulated rules, not with respect to a historically given natural language. (Carnap, 1952b, 432)[4]

In my view, it is crucial to treat the case of analyticity in formal languages separately because formal languages are importantly different from natural ones. I will substantiate

[1]See for instance Boghossian (1996, 2003a), Peacocke (1992), Russell (2008), and Williamson (2006, 2007).

[2]A sentence stating a rule of inference for a formal language is a sentence of the metalanguage. For instance, the sentence 'B may be inferred from $\neg A \vee B$ and A,' which states the rule of inference of disjunctive syllogism, is part of the metalanguage (here English), and 'A' and 'B' are syntactic variables in the metalanguage that range over formulas of the formal language in question. Strictly speaking, '$\neg A$' and '$A \vee B$' should be written respectively as '$\neg$' $^\frown A$ and $A^\frown$'$\vee$'$^\frown B$, but I will use the former shorthands throughout.

[3]See for instance Ayer (1936), Carnap (1937), and Hahn (1959).

[4]See also: "It must be emphasized that the concept of analyticity has an exact definition only in the case of a language system, namely a system of semantical rules, not in the case of an ordinary language, because in the latter the words have no clearly defined meaning" (Carnap, 1952b, 427). A. J. Ayer notes in a similar vein, as he explains how to understand 'truth in virtue of meaning': "There is ground for saying that the philosopher is always concerned with an artificial language. For the conventions which we follow in our actual usage of words are not altogether systematic and precise" (Ayer, 1936, 70, fn. 1).

this point by looking at the most powerful contemporary argument against analyticity, due to Timothy Williamson (2006; 2007). Williamson first makes an argument aimed at showing that there is no analyticity in natural language, and then extends it to formal languages. I will argue that this extension doesn't work. This will enable me to side with the logical empiricists and recognize a notion of analyticity whose natural home, so to speak, is in formal mathematical languages. I call this notion of analyticity 'formal analyticity'.

Recognizing formal analyticity has at least two payoffs. First, formal analyticity may subserve certain views in the epistemology of logic and mathematics. Traditional notions of analyticity, as well as many contemporary ones, were purported to have epistemological consequences: The claim that some sentence is analytic was supposed to imply that it is knowable (or that one can have justification for it) merely on the basis of possessing certain concepts or understanding certain words.[5] Although epistemological consequences of this strength will not follow from formal analyticity, it will reveal that our knowledge of some mathematical axioms and rules of inference crucially involves concept possession or linguistic understanding. Second, even though understanding a formal language is fundamentally different from understanding a natural language, certain fragments of natural language may have enough similarity with formal languages that formal analyticity can shed light on them. The tools developed in this paper should

[5] See for instance Ayer (1936), Hahn (1959), and more recently Bealer (1999), Boghossian (1996, 2003a), and Peacocke (1992).

prove useful in seeking an analogue to formal analyticity in natural languages.

Here is the plan. In Section 2.1, I will explain Williamson's argument against analyticity and how it is supposed to apply to formal languages. I will argue that Williamson's argument, when applied to formal languages, is ambiguous between an "external" and an "internal" reading. On the external reading, which I will examine in Section 2.2, the argument is sound but committed to an implausible view of what it is to understand a formal language. On the internal reading, which I will examine in Section 2.3, the argument is unsound. The internal reading of Williamson's argument will provide the material to define the notion of formal analyticity, which I will explain and defend in Section 2.4. I will conclude, in Section 2.5, by saying more about the potential payoffs of formal analyticity.

2.1 Williamson's argument

In (2006; 2007), Williamson claims the following argument template can be used to show that any given sentence, s, is not analytic:

1. If s is analytic, then, necessarily, whoever understands s assents to s. [*premise, from definition of 'analytic'*[6]]
2. It is possible that someone understands s but doesn't assent to s. [*premise, with independent argument*]
3. s is not analytic. [*conclusion, from 1 and 2*]

[6] Williamson takes this to be a definition of the "epistemological conception of analyticity" (Williamson, 2007, 73). Notice that Williamson doesn't need to accept any definition of 'analytic'; premise 1 (and whatever definition of 'analytic' underlies it) is understood to be proposed by Williamson's opponents.

For Williamson, assent is a dispositional mental attitude, which implies that one can assent without actively or overtly assenting (Williamson, 2007, 74).[7] Williamson suggests that one can also replace 'assent' with 'belief' throughout the argument, where 'belief' may stand either for implicit or explicit belief (Williamson, 2007, 74). Understanding a sentence, in turn, consists in understanding its constituent words and syntax, and is a matter of having "normal linguistic competence" (Williamson, 2007, 74).

Williamson intends his argument template to work also against the claim that certain rules of inference are analytic in the sense that assenting to them is necessary for understanding the relevant logical constants (Williamson, 2007, 76). According to Williamson, assent to a rule of inference is to be understood as assent to argument instances licensed by that rule of inference (Williamson, 2007, 75f., 92). Since it is commonly thought that the bearers of analyticity are sentences, and since it is unclear whether rules of inferences are sentences, let us take 'rule of inference' to be interchangeable with 'sentence stating a rule of inference', and restate Williamson's claim as follows. Let s in the argument template above also range over sentences stating rules of inferences for logical constants. Then assume that understanding a sentence stating a rule of inference for a logical constant is primarily a matter of understanding that logical constant, and take assent to a sentence stating a rule of inference to be assent to argument instances

[7] Later Williamson goes on to argue against "watered down" understanding–assent links with 'assent' replaced throughout with 'disposition to assent' (Williamson, 2007, 100–109). He states that "[h]aving a disposition to assent does not entail assenting" (Williamson, 2007, 100). There are thus two notions of "disposition to assent" at play for Williamson; here I will remain neutral between the two, as this won't make a difference to my arguments.

licensed by that rule of inference.

Williamson extends his argument to formal languages: He claims his argument template works even for sentences or rules of inference of "carefully constructed formal languages" (Williamson, 2007, 94). Before we examine his argument for this, let us first get clear on the stakes of Williamson's argument against analyticity.

According to Williamson, analyticity places "bedrock or constitutive" constraints on understanding certain sentences: To say that a sentence is analytic is to say something about what it is to understand that sentence (Williamson, 2006, 4). He later diagnoses the problem with the claim that some sentence is analytic as "depend[ing] on a misapprehension of what it is to possess a concept or to understand a word" (Williamson, 2006, 8). So what is at stake in Williamson's argument is an account of linguistic understanding (or concept possession). For Williamson's opponents, linguistic understanding, at least sometimes, goes along with certain "fixed-points": For some words, w, there is a collection of sentences, Π_w, to which one needs to assent in order to be competent with w. Analytic sentences, on this view, are those sentences in Π_w that contain w: Understanding them requires assenting to them.[8] Williamson's view on linguistic competence rejects that such sets of fixed-points Π_w exist for any given word w. He proposes that understanding is compatible with the absence of any "common creed" whatsoever

[8]We may want to leave it open whether all sentences in Π_w are analytic (i.e. such that understanding them requires assenting to them). Indeed, it may be that for some words, w, understanding w requires assenting to some sentence, s, which doesn't contain w, and such that for any word, w', in s, $s \notin \Pi_{w'}$. In this case (also assuming that understanding the syntax of s doesn't require assenting to s), understanding s wouldn't require assenting to s, even though $s \in \Pi_w$.

(Williamson, 2007, 125), and instead requires standing in appropriate (mainly causal) relations to the community of speakers:

> What binds together uses of a word by different agents or at different times into a common practice of using that word with a given meaning …is the complex interrelations of the constituents, above all, their causal interrelations. (Williamson, 2007, 123)
>
> One can lack understanding of a word through lack of causal interaction with the social practice of using that word, or through interaction too superficial to permit sufficiently fluent engagement in the practice. But sufficiently fluent engagement in the practice can take many forms, which have no single core of agreement. (Williamson, 2007, 126)

Williamson's account of linguistic understanding is closely related to externalist accounts introduced by Hilary Putnam (1975b) and Tyler Burge (1979). According to externalism, an individual can understand a word even while being ignorant or wrong about the application conditions of that word in her linguistic community, because she can, for instance, defer to others in her linguistic community. In traditional externalist arguments, however, deferential understanding relies on the existence of some people in the linguistic community (the experts) who know the application conditions of the word and who thus have complete understanding of the word. A natural way to reply to externalism for a defender of analyticity is thus to restrict her claims to complete understanding: For some words, w, there is a collection of sentences, Π_w, to which one needs to assent in order to completely understand w. Whether or not this reply settles the issue, it is hard to make against Williamson's argument. Indeed, the crucial difference between Williamson's account of understanding and traditional externalist accounts is that linguistic competence requires no such fixed-points Π_w even among experts with "full understanding" of w (Williamson, 2007, 98). This view is supported by his argument against analyticity: Most of Williamson's examples in support of premise 2 are hypothetical or actual experts about the subject matter of the candidate analytic sentence.[9] To resist Williamson's argument by upholding understanding–assent links for subjects with complete understanding, the defender of analyticity would thus have to construe the notion of complete understanding such that even the relevant experts might lack complete understanding. But it is hard to see how this notion would be independently motivated, and how it would fit with the "ordinary" notion of understanding with which Williamson is concerned.[10]

My arguments in this paper will support the claim that for formal mathematical languages, linguistic understanding—or, at the very least, complete linguistic understanding—does require a common creed. So what is primarily at stake in the present paper is an account of understanding of formal mathematical languages.

Let me first make a clarification about how to formulate the claim that some sentences of formal languages are analytic. Analyticity is a relation between a sentence and its language, and a language is more than a mere set of strings of symbols: It is a set of strings of symbols with their respective meanings. Meaning here is understood, roughly, as something that characterizes truth or truth in a structure, and thus determines in one

[9]See for instance Williamson (2007, 123f., 125).

[10]See for instance Williamson (2007, 74). George Bealer's notion of understanding might be an example of such a notion of complete understanding (Bealer, 1999).

or another way a consequence relation. Accordingly, if a sentence of a formal language is going to count as analytic, then we should not think of its analyticity as a relation merely to a formal language in the strict sense (i.e. a set of strings of symbols specified by giving an alphabet and formation rules) but as a relation to a formal language in the sense of either (i) a formal system (a set of strings of symbols conjoined with derivation rules, i.e. axioms and rules of inference), or (ii) a formal system conjoined with a formal semantics. The formal semantics associated with a formal system at least specifies the available truth-values or degrees (e.g. $\{\top, \bot\}, \{\top, \bot, \mathbb{I}\}$, $[0, 1]$) and assigns logical connectives ('$\vee$', '$\neg$') to truth-functions. Until Section 2.4, I will only be concerned with option (ii). In Section 2.4, I will propose a definition of analyticity for each of these two options: Syntactic Formal Analyticity goes along with option (i), while Semantic Formal Analyticity goes along with option (ii).

Williamson extends his argument against analyticity to formal languages as follows. He uses his argument template with disjunctive syllogism, i.e. the rule of inference stating that B may be inferred from A and $\neg A \vee B$. In support of the second premise of that argument, Williamson points to Graham Priest. Priest is a philosopher and logician who, according to Williamson, is linguistically competent with '$\vee$', '$\neg$' if anyone is and hence understands disjunctive syllogism, but doesn't assent to disjunctive syllogism. Williamson states that this is because, as a dialetheist, Priest believes that some sentence A can be both true and false, in which case, if B is only false, the premise $\neg A \vee B$ of disjunctive syllogism is true, while the conclusion B is only false (Williamson, 2007, 94).[11]

Priest is given as a knockdown counterexample to the analyticity of disjunctive syllogism, but his case is importantly underdescribed. In particular, there is an ambiguity in what 'assent' to a formula or rule of inference of a formal language means. If we look at the way in which Priest "doesn't assent" to disjunctive syllogism, we get an "external" reading of 'assent': 'Assent to *s*' when *s* is a formula (or rule of inference) of a formal language means assent to *s* being true (or truth-preserving) "in reality," in some sense I will make precise. But there is also a weaker and more natural, "internal" reading of 'assent', according to which 'assent to *s*' means assent to *s* being true (or truth-preserving) according to the formal semantics of a given formal system, without any further commitment to *s*'s being true "in reality." In what follows, I will explain each reading and evaluate the two arguments against analyticity that result from them in turn.

2.2 The external account of 'assent'

Throughout works such as Priest (1979, 1995), Priest uses formal systems such as the Logic of Paradox (henceforth 'LP') to "model," "capture," or "represent" certain facts.

[11] Williamson also raises the same point more recently, this time by using the example of '∈' in the formal language of set theory: "Some philosophers think that the key to the epistemology of first principles of logic is that they are *analytic*, in the sense that a disposition to assent to them is essential to linguistic or conceptual competence with the logical constants that figure in them. ...A well-trained set theorist may come to challenge one of the axioms of ZFC on theoretical grounds, and cease to find it primitively compelling, without incurring any credible charge of linguistic or conceptual incompetence with the word 'set' or the '∈' symbol for set membership" (Williamson, 2016, 247f.). I will examine the case of the language of set theory in Section 2.4.

Here in Priest (1979), he is mainly interested in representing ordinary mathematical reasoning carried out in natural language:

> The formal logician is essentially an applied mathematician. It is his job to construct mathematical systems which model (in the physicist's sense, not the logician's) some natural phenomenon. The phenomenon the logician is particularly interested in, is normal (naïve) reasoning carried out in a natural language. …And the mathematical systems he uses are formal languages, mathematical semantics, etc. (Priest, 1979, 225)

Priest's famous claim that there are true contradictions (i.e. dialetheism), on the other hand, seems to point to his interest in representing "logical reality":

> How could a contradiction be true? After all, orthodox logic assures us that for every statement, α, only one of α and $\neg\alpha$ is true. The simple answer is that orthodox logic, however well entrenched, is just a theory of how logical particles, like negation, work; and there is no a priori guarantee that it is correct. (Priest, 1995, 4)

For now, let us focus on Priest (1979). There, Priest argues that in part because we can "naïvely but not formally prove" that the Gödel sentence is true, only formal systems that contain their own truth predicate can adequately represent ordinary mathematical reasoning (Priest, 1979, 224). It is for this purpose that he puts forward LP, which is a semantically closed, paraconsistent, and non-classical formal system. Right after explaining LP, Priest mentions that disjunctive syllogism is invalid according to the formal semantics of LP, and states that this is actually a problem for his project:

> [T]he aim of the exercise was to construct a logic which could be used (in connection with a semantically closed theory) to capture naïve mathematical reasoning. However, eschewing [disjunctive syllogism] …would obviously have a crippling effect on mathematical reasoning. (Priest, 1979, 232)

The interest of this passage for our purposes is that it shows Priest's grievance isn't specifically with disjunctive syllogism, or with any other particular rule of inference, but rather with entire formal systems; Priest seeks a formal system that, as a whole, correctly represents ordinary mathematical reasoning carried out in natural language, and he argues that classical formal systems cannot fulfill this role.

So we have reason to think that, at least in Priest (1979), Priest "doesn't assent" to disjunctive syllogism in the sense that he doesn't think that disjunctive syllogism is part of a formal system that correctly represents ordinary mathematical reasoning carried out in natural language. We can now generalize this understanding of 'assent' to include the other uses Priest makes of formal systems, and state our "external" account of 'assent': To assent to s, where s is a formula (or rule of inference) of a formal system, is to assent to s being part of a formal system that correctly represents Reality, where 'Reality' henceforth may stand either for "logical reality" or ordinary (mathematical) reasoning carried out in natural language. This, in turn, yields the following "external" interpretation of Williamson's argument:

1e. If disjunctive syllogism is analytic, then, necessarily, whoever understands disjunctive syllogism assents to disjunctive syllogism's being part of a formal system that correctly represents Reality. [*premise, from definition of 'analytic'*]

2e. Graham Priest understands disjunctive syllogism but doesn't assent to disjunctive syllogism's being part of a formal system that correctly represents Reality. [*premise*]

3e. Disjunctive syllogism is not analytic. [*conclusion, from 1e and 2e*]

The foregoing discussion served to show that premise 2e is plausible. I now want to argue against 1e—or, more precisely, against the definition of 'analytic' that must underlie 1e. I will consider two candidate definitions of 'analytic' that can underlie 1e and raise some problems for each.

2.2.1 First candidate external definition of 'analytic'

One might think that "assenting to disjunctive syllogism's being part of a formal system, Σ, that correctly represents Reality" implies having the belief that Σ adequately represents Reality. If so, and if one wants to claim that disjunctive syllogism is analytic in, say, First-Order Logic (henceforth 'FOL'), then 1e—and, a fortiori, whatever definition of analyticity underlies 1e—would imply that whoever understands disjunctive syllogism believes that FOL correctly represents Reality. Let me henceforth use 'a sentence of a formal system' to refer either to a formula or rule of inference of a formal system. Here is the simplest candidate definition of 'analytic' that would have such an implication:

1e′. A sentence, s, of a formal system, Σ, is *analytic* in Σ just in case, necessarily, whoever understands s believes that Σ correctly represents Reality and that s is part of Σ. [*definition*][12]

Here, 's is part of Σ' could for instance mean that s is a derivation rule, a theorem, or true in the standard model of Σ; I will leave it open how exactly to interpret this.

[12] One could also propose a weaker definition 1e′−: A sentence, s, of a formal system, Σ, is *analytic* in Σ just in case, necessarily, whoever understands s believes that s is part of a formal system that correctly represents Reality. My arguments against 1e′ below also work against 1e′−, so I won't consider 1e′− separately.

The problem with 1e′ is that if there are analytic sentences of formal systems, then 1e′ would place unreasonable demands on what counts as being linguistically competent with them. According to 1e′, one is not linguistically competent with some sentences of a formal system unless one believes that the formal system correctly represents some aspect of Reality. But this is an unreasonable constraint on linguistic understanding in formal systems: One needn't even ever have had beliefs about "logical reality," "mathematical reality," or ordinary (mathematical) reasoning to be linguistically competent in formal systems.

Let me illustrate this point with an example. On a natural interpretation, Carnap didn't believe that his formal systems adequately represent "logical reality" in Priest's sense. In (1950a), Carnap states that accepting a Tarskian truth-definition for a formal system doesn't commit one either way on the "pseudo" question concerning whether mathematical entities really exist:

> But it is certainly wrong to regard my semantical method [i.e. inductively defining "truth"] as involving a belief in the reality of abstract entities, since I reject a thesis of this kind as a metaphysical pseudo-statement. (Carnap, 1950a, 36)

On a natural interpretation, Carnap also didn't believe that his formal systems correctly represent ordinary (mathematical) reasoning carried out in natural language. Carnap thought of the process of formalization as an "explication." An explication, for Carnap, takes an inexact (ordinary) concept and replaces it with an exact (formal) one defined in a well-constructed formal system (Carnap, 1950b, 3). Although the exact concept should overlap to some extent with the informal one, Carnap notes that "close similarity is not required, and considerable differences are permitted" (Carnap, 1950b, 7), because for him the purpose of the formal system is to be scientifically "fruitful," not to "correctly represent" the informal notion.

On a natural interpretation, then, Carnap doesn't believe that the formal systems he studied correctly represent Reality in our sense above. If 1e′ is right, then people like Carnap who don't believe that a given formal system correctly represents Reality, or who simply don't have beliefs about these issues, wouldn't count as understanding analytic sentences of that formal system. These people include many defenders of analyticity and many mathematicians who have a formalist philosophical conception of their fields.[13] This shows that there are no analytic sentences of formal systems, if 'analytic' is understood as in 1e′. But it also suggests that it is uncharitable to ascribe 1e′ as a definition of analyticity to the defenders of analyticity in formal systems.

2.2.2 Second candidate external definition of 'analytic'

There might be a way to interpret 1e such that someone who assents to disjunctive syllogism's being part of a formal system that correctly represents Reality doesn't need to have beliefs about the formal system or Reality. Perhaps all that is needed for assenting in this way is to have one's behavior, beliefs, or reasoning "in Reality" accord with the principles of the formal system. More precisely, in the case of disjunctive syllogism,

[13] A famous example of a "formalist" mathematician is Paul Cohen. See for instance Cohen (1971).

the idea would be that if disjunctive syllogism is analytic in FOL, then understanding disjunctive syllogism requires assenting to arguments by disjunctive syllogism when one ordinarily reasons using natural language. Generalizing this idea yields the following definition of analyticity:

1e″. A sentence, s, of a formal system, Σ, is *analytic* in Σ just in case, necessarily, whoever understands s assents to s in Reality. [*definition*]

Here, 'assents to s in Reality' is a shorthand for 'assents to s (or the natural language translation of s) when they ordinarily reason using natural language', or 'assents to s (or the natural language translation of s) when they carry out naïve mathematical reasoning'.

If 1e″ is the definition of 'analytic' that underlies the first premise of Williamson's argument when it is applied to formal systems, then this argument collapses into his argument against analyticity in natural languages. Right before introducing Priest's example, Williamson cites Vann McGee in support of the claim that some people understand "if … then" in natural language, but don't assent to arguments by modus ponens. The example of Priest, then, comes in support of the claim that "[t]he problem is not just the vagueness of natural languages. Similar problems arise for carefully constructed formal languages" (Williamson, 2007, 94). But now, 1e″ states that understanding an analytic sentence of a formal system Σ requires assenting to it (or to some translation of it) in natural language. So, contrary to Williamson's claim, we lose the distinction between analyticity for formal languages and analyticity for natural languages.

If $1e'$ and $1e''$ were the only available understandings of analyticity for formal languages, then Williamson would be right that there are no analytic sentences of formal languages. However, I think the current reading of Williamson's argument overlooks a natural understanding of 'assent' when it is applied to sentences of formal languages—one that can yield a notion of analyticity for formal languages. In what follows, I argue that there is such a notion of analyticity—formal analyticity—and that Williamson's argument template doesn't work against it. This will supplement my arguments here in Section 2.2 that it would be uncharitable to ascribe an external definition of analyticity (either 1e$'$ or 1e$''$) to defenders of analyticity for formal languages.

2.3 The internal account of 'assent'

In Section 2.2, I examined the sense in which Priest "doesn't assent" to disjunctive syllogism. Here in Section 2.3, I will examine a weaker and more natural sense of 'assent' to sentences of formal systems.

As I stated in the introduction, for Williamson, failure to assent to a rule of inference in natural language is manifested in failure to assent to some argument instance(s) licensed by that rule of inference. The latter, in turn, is understood as assent to premises but not to conclusions of argument(s) licensed by that rule of inference (Williamson, 2007, 92). Let us now turn to formal systems. When we talk about failure to assent to the rule of inference disjunctive syllogism of a formal system, what should we mean? By analogy, we should say that failure to assent to the rule of inference disjunctive syllogism of a formal system is manifested in failure to assent to conclusions while assenting to premises of some argument instances licensed by disjunctive syllogism. Premises and conclusions of an argument in a formal language are formulas of the formal language of the form $A \vee B$ and $\neg A$, where 'A' and 'B' are syntactic variables of the metalanguage ranging over formulas of the formal language. We can now ask the same question at the level of formulas: When we talk about assent to a formula of a formal system, what should we mean? On a natural proposal, to assent to a formula of a formal system is to assign the truth-value $\top$ to the formula, i.e. to assign a formal interpretation to the formula under which it gets the truth-value $\top$.

But notice that assigning such a formal interpretation to a formula presupposes a set of available truth-values, and an interpretation of the logical connectives (i.e. an assignment of truth-functions). Different formal systems give different interpretations of the logical connectives; for instance, FOL assigns the classical truth-functions to '$\neg$' and '$\vee$', while LP assigns non-classical, three-valued truth-functions. So on this second account of 'assent' to formulas of formal systems, which I shall call 'the internal account', assent is relative to some given formal system and its associated semantics. A fortiori, assent to rules of inference of a formal system is also relative to a given formal system and semantics. I will say 'assent to disjunctive syllogism in Σ' to talk about this internal kind of assent to disjunctive syllogism relative to formal system Σ and its semantics.

Disjunctive syllogism is truth-preserving according to the semantics of FOL, while it isn't truth-preserving according to the semantics of LP—where a rule of inference is truth-preserving in some formal system, Σ, if there is no assignment of truth-values that assigns $\top$ to the premises and also $\bot$ to the conclusion of an argument by that rule of inference (where the truth-values of logically complex formulas are computed according to the truth-functions assigned to the logical connectives by the formal semantics of Σ). Anyone who knows these should, a fortiori, assent to disjunctive syllogism in FOL and not assent to disjunctive syllogism in LP. From this, we get two interpretations of Williamson's argument template under the internal account of 'assent' in formal languages. Here is the first:

1i. If disjunctive syllogism is analytic in FOL, then, necessarily, whoever understands disjunctive syllogism assents to disjunctive syllogism in FOL. [*from definition of 'analytic'*]

2i. Graham Priest understands disjunctive syllogism but doesn't assent to disjunctive syllogism in FOL. [*premise*]

3i. Disjunctive syllogism isn't analytic in FOL. [*conclusion, from 1i and 2i*]

But 2i is false, as Priest knows that disjunctive syllogism is truth-preserving in FOL and hence also assents to disjunctive syllogism in FOL.

Here is the second interpretation:

1i′. If disjunctive syllogism is analytic in LP, then, necessarily, whoever understands disjunctive syllogism assents to disjunctive syllogism in LP. [*from definition of 'analytic'*]

2i′. Graham Priest understands disjunctive syllogism but doesn't assent to disjunctive syllogism in LP. [*premise*]

3i′. Disjunctive syllogism isn't analytic in LP. [*conclusion, from 1i′ and 2i′*]

Here, a defender of analyticity should agree with 3i′. No sensible defender of analyticity would propose that disjunctive syllogism is analytic in LP, given that disjunctive syllogism is not even truth-preserving in LP. So, instead of endorsing 1i′ and its antecedent (i.e. that disjunctive syllogism is analytic in LP), a defender of analyticity should endorse 1i and its antecedent (i.e. that disjunctive syllogism is analytic in FOL). But, as we have just seen, Williamson doesn't raise a problem for this, as he doesn't provide an example of someone who understands disjunctive syllogism and yet fails to assent to it in FOL.

In what follows, I will argue that this isn't just an easily remedied oversight. Williamson *cannot* provide such an example: There cannot be examples of people who understand disjunctive syllogism and yet fail to assent to it in FOL. More generally, there is a plausible and non-empty definition of analyticity for formal systems that underlies 1i.

2.4 Formal Analyticity

Consider the following definitions:

Semantic Formal Analyticity A formula (or rule of inference), s, of a formal system, Σ, is *semantically formally analytic* in Σ just in case, necessarily, whoever understands s believes that s is true (or truth-preserving) according to the formal semantics of Σ.

Syntactic Formal Analyticity A formula (or rule of inference), s, of a formal system, Σ, is *syntactically formally analytic* in Σ just in case, necessarily, whoever understands s believes that s is derivable in (or is a rule of inference of) Σ.[14]

Given the internal reading of 'assent', Semantic Formal Analyticity instantiated with disjunctive syllogism implies 1i. Syntactic Formal Analyticity, on the other hand, is a reasonable alternative definition of analyticity for formal systems that is not based on semantic notions (cf. option (ii) in Section 2.1). Let me make one clarification concerning Syntactic Formal Analyticity: I will assume that if s is an axiom of the formal system Σ, then either believing that s is an axiom of Σ or believing that s can be used as an assumption in derivations in Σ are sufficient for believing that s is derivable in Σ.[15]

The task at hand is to argue that there are formally analytic sentences. In what follows, I will only consider two cases, one for each kind of formal analyticity: the rule of inference of disjunctive syllogism in FOL as a case of semantic formal analyticity, and the Extensionality axiom of Zermelo–Fraenkel Set Theory with Choice (henceforth 'ZFC')

[14]Of note is that these definitions are very similar to the ones proposed by Carnap (1937; 1939). For Carnap, analytic sentences of a mathematical formal system are (roughly) all those that are true according to the formal semantics of that formal system (Carnap, 1937, 111). Analyticity for Carnap is thus a purely formal feature of a sentence of a formal system; it doesn't make reference to the notion of linguistic understanding because Carnap found this notion is too imprecise (Carnap, 1939, 12f.). However, Carnap agrees with the idea motivating formal analyticity that understanding formal languages has to do with knowing their derivation rules: "Since to know the truth conditions of a sentence is to know what is asserted by it, the given semantical rules determine for every sentence of [the formal system] …what it asserts—in usual terms, its "meaning"—or, in other words, how it is to be translated into English. …Therefore, we shall say that we *understand* a language system, or a sign, or an expression, or a sentence in a language system, if we know the semantical rules of the system" (Carnap, 1939, 10f.).

[15]In general, a formal derivation in Σ is a sequence of formulas such that each formula in the sequence is either an axiom, or else results by a rule of inference from formulas that precede it in the sequence. So axioms of a formal system are trivially derivable in that formal system. I make this assumption here concerning Syntactic Formal Analyticity because I want to make room for a (possibly) weaker reading.

as a case of syntactic formal analyticity. The case of ZFC doesn't arise in Williamson's arguments, but it is of contemporary interest: Influenced by logical empiricists, Kurt Gödel famously claimed that some axioms of set theory are analytic in that they "only unfold the content of the concept of *set*" (Gödel, 1964, 477), and this claim has recently gained attention among mathematicians and philosophers who discuss so-called 'intrinsic justifications' for new axioms of set theory.[16] Formal analyticity is most likely not the notion of analyticity Gödel had in mind, but, as we will see in Section 2.5, it may still shed light on the issue of intrinsic justification.

I first want to argue that there are no experts who can count as counterexamples to the semantic formal analyticity of disjunctive syllogism in FOL, or to the syntactic formal analyticity of Extensionality in ZFC. In other words, I claim that if someone is an expert concerning propositions of FOL, then, necessarily, if they understand disjunctive syllogism (or '$\vee$'), they believe that disjunctive syllogism is truth-preserving according to the formal semantics of FOL. Similarly, if someone is an expert concerning propositions of ZFC, then, necessarily, if they understand Extensionality (or '$\in$'), they believe that Extensionality is derivable in ZFC. I begin by considering the case of experts because, as we have seen in Section 2.1, all of Williamson counterexamples to analyticity are experts in the domain of the putative analytic sentences. The considerations used to show that there are no expert counterexamples to formal analyticity will then help us

[16]See for instance Boolos (1971), Eagle (2008), Feferman (1999), Koellner (2009), Maddy (1988, 2011), Martin (ms), and Welch (ms).

mount an argument for the existence of formal analyticity more generally.

Consider the case of Extensionality. Extensionality is one of the most basic axioms of ZFC; it states that sets are identical just in case they have the same members.[17] Now the experts concerning propositions of ZFC are presumably the set theorists. And every set theorist knows that Extensionality is an axiom of ZFC (and also that Extensionality can be used as an assumption in derivations in ZFC, and that Extensionality is derivable in ZFC). After all, set theorists are people who work with ZFC and on the properties of ZFC; they must at least know what are the axioms of ZFC in order to do their work.

In general, I think it is very plausible that being an expert concerning propositions of ZFC requires knowing at least what are the basic axioms of ZFC;[18] it would just be very odd to call someone such an 'expert' who didn't know what are the basic axioms of ZFC. But I propose we can also provide an argument for this, with the following two premises. First, if someone is an expert concerning propositions of ZFC, then they are able to do basic derivations in ZFC. Second, in order to be able to do basic derivations in ZFC, one needs to know what are the basic axioms of ZFC. So experts concerning ZFC need to know (and thus believe) of the basic axioms of ZFC that they are axioms of ZFC. Let us consider these two premises in turn.

An expert concerning propositions of a formal language is someone who is able to

[17] I.e. $\forall x, y(x = y \leftrightarrow \forall z(z \in x \leftrightarrow z \in y))$.

[18] Throughout, I will understand "knowing what are the basic axioms of ZFC" as knowing *of* the basic axioms of ZFC that they are axioms of ZFC, without necessarily knowing or believing that these axioms are *basic*.

use that language (or the expressions of the language). Using a formal language (or an expression of a formal language), in turn, at the very least involves doing some basic derivations in the formal language.[19] After all, there is no other use of formal languages; formal languages are artificial devices used exclusively in the context of doing mathematics. And using a formal language in the context of doing mathematics at the very least involves doing some basic derivations in the formal language.

Now consider the second premise. A formal derivation in ZFC is a sequence of formulas such that each formula in the sequence is either an axiom, or else results by a rule of inference from formulas that precede it in the sequence. So, being able to do formal derivations in ZFC requires being able to put together (in writing or in thought) a sequence of formulas of ZFC that includes some axioms of ZFC. The most basic formal derivations are going to use some of the most basic axioms of the formal system, like Extensionality, Foundation, or Pairing (in fact, most derivations will require Extensionality). So, being able to do basic formal derivations in ZFC at least requires knowing of the most basic axioms of ZFC that they are axioms of ZFC. Being able to do basic formal derivations in ZFC thus requires believing that Extensionality is an axiom of ZFC (or treating Extensionality as an axiom of ZFC).

I just argued that there won't be experts who count as counterexamples to the syntactic formal analyticity of Extensionality in ZFC. A similar argument can be made

[19] In order to count as an expert, one has to be able to use the language in a way that exhibits understanding. And thus, if one of the basic uses of the language is to do formal derivations, this ability also needs to involve understanding, as opposed to mere manipulation of symbols.

for why there won't be experts who count as counterexamples to the semantic formal analyticity of disjunctive syllogism in FOL. First, if someone is an expert concerning propositions of FOL, then they should be able to do basic semantic reasoning in FOL (e.g. they should be able to use truth-tables). Second, in order to be able to do basic semantic reasoning in FOL, one needs to know what are the basic semantic facts about FOL, including the fact that disjunctive syllogism is truth-preserving in FOL. More generally, the argument I propose for why there won't be expert counterexamples to the formal analyticity of the most basic axioms and rules of inferences of formal systems is that (Premise 1) being an expert with respect to a given formal system requires being able to do some basic mathematics with the formal system, and (Premise 2) being able to do some basic mathematics with a formal system requires knowing of the basic axioms, rules of inference, and formal semantics of that formal system that they are axioms, rules of inference, and formal semantics of that formal system. So one cannot find experts who provide counterexamples to the formal analyticity of at least the most basic axioms and rules of inferences of a given formal system.

At this point, we have already undermined Williamson's argument against analyticity in formal languages. Williamson had claimed that there is no sentence, s, of a formal language, Σ, such that all experts in Σ assent to s; we have just argued against this. Williamson's argument against analyticity is thus not distinctively powerful in having all counterexamples to analyticity be hypothetical or actual experts (cf. Section 2.1). Moreover, Williamson's argument against analyticity in natural languages doesn't extend to formal languages because there are no experts who provide counterexamples to the understanding–assent links in formal analyticity. So we have achieved one of the aims stated at the outset of the paper. This doesn't yet get us to the claim that there are formally analytic sentences, however: All experts might know that Extensionality is derivable in ZFC, but their knowledge might go beyond linguistic understanding; linguistic understanding of ZFC alone might not require knowing that Extensionality is derivable in ZFC. We thus need an independent argument for the claim that there are formally analytic sentences like Extensionality or disjunctive syllogism.

My argument will be based on the considerations used above to show that there won't be expert counterexamples to formal analyticity. More specifically, I will argue that the basic expert ability I identified above is also a requirement for linguistic understanding of the formal languages at issue. The argument turns on a constraint on linguistic understanding. This constraint is theory neutral—it is implied by our ordinary practice of ascribing linguistic or conceptual competence, not by a particular account of concept possession or linguistic understanding—and it is widely accepted. Many philosophers, including paradigm externalists like Burge, Putnam, and Williamson himself, accept it. I will argue that this constraint is compatible with the claim that there are no analytic sentences in natural languages, but that it implies that there are formally analytic sentences.

Consider the constraint:

Understanding–Ability Understanding an expression, e, requires having the ability to use e in basic ways.

What are the "basic ways" of using a given expression, e? This will depend on what e is, and on what is the role of e in the linguistic community. For instance, the word 'vixen' is typically used to single out certain kinds of animals (vixens). So the basic ways of using 'vixen' will, presumably, include applying 'vixen' correctly at least sometimes, and not applying 'vixen' consistently and only to ironing boards, for instance. In general, it is plausible that an expression's normal context of use determines what are the basic ways of using it.

Putnam, Burge, and Williamson himself endorse Understanding–Ability.[20] According to Putnam, for instance, understanding an expression requires being able to use it in a way that "passes muster," which means, for Putnam, that speakers must at least be able to apply the words correctly to some basic cases (for instance, not apply 'tiger' to a snowball (Putnam, 1975a, 168)).

Similarly, according to Burge, understanding a concept requires having at least very basic discriminatory abilities:

> Having …[the concept of arthritis] requires having certain associated discriminatory abilities. …Thus the individual must be able to discriminate arthritis from such things as animals, trees, and numbers, and from certain

[20]Other contemporary thinkers also endorse Understanding–Ability, albeit in different contexts. For instance, John Bengson and Mark A. Moffett (2007) argue that some concepts are "ability-based concepts," i.e. concepts understanding of which entails having certain abilities (see also Bengson (2016) and Pavese (2015)).

> other diseases, in order to have the concept. But he need not be able to discriminate it from all other rheumatoidal diseases, actual or possible—except insofar as he does so by employing the concept *arthritis*. (Burge, 1979, 325)

Williamson also commits himself to Understanding–Ability. For instance, he states that "[u]nderstanding words in a natural language has much to do with the ability to use them in ways that facilitate smooth and fruitful interactions with other members of the community" (Williamson, 2007, 97).[21] There is, of course, room to discuss which are the abilities to use w in "basic ways," but, on a natural reading, all of Putnam, Burge, and Williamson accept Understanding–Ability in its general form.

There is, however, an important point of divergence between Putnam and Burge on the one hand, and Williamson on the other. On a natural reading, Putnam and Burge are also committed to understanding–assent links, i.e. to the existence of analytic sentences in Williamson's sense. Indeed, it is natural to read Putnam as claiming that there are fixed-points for understanding: Understanding an expression requires knowing what is the stereotype typically associated with the expression in one's community. For instance, Putnam claims that in our linguistic community, understanding 'tiger' at least requires knowing what tigers look like (Putnam, 1975a, 168).[22] On a natural reading of

[21] Williamson also clearly assumes something like Understanding–Ability throughout his arguments against analyticity. For example, in making the case that his hypothetical example, Peter, understands the sentence 'Vixen are vixen' but doesn't assent to it, Williamson states that Peter is able to have normal conversations about vixens with members of his linguistic community, that he is able to adjust his conversations so that his odd beliefs about vixens don't get him in trouble, etc. (Williamson, 2007, 88–91).

[22] See also Putnam (1970).

Burge, in turn, he also holds that there are fixed-points for complete understanding: For instance, the people who completely understand 'arthritis' (the doctors) should know that arthritis is a disease of the joints (Burge, 1979, 84ff.). Williamson argues specifically against Putnam's proposal (2007, 109f.), and his response to Burge-style proposals is, as we have seen in Section 2.1, to restrict all his counterexamples to experts who are purported to have complete understanding.

More generally, Williamson's arguments in (2006; 2007) can be understood as arguments for the claim that Understanding–Ability doesn't imply understanding–assent links. As we saw in Section 2.1, Williamson claims that understanding an expression has to do with "sufficiently fluent engagement in the practice" of using that expression (Williamson, 2007, 126), but that this need not require a "single core of agreement" (Williamson, 2007, 126). Williamson claims that "[s]peakers may simply differ from each other in various ways in their ability to distinguish between members and non-members of the relevant kind" (Williamson, 2007, 124).[23] In other words, there is no fixed set of sentences to which one needs to assent in order to possess the ability to use the word in basic ways. One can have the ability to use 'vixen' in basic ways by believing that vixens are female foxes, or by believing that vixens are not foxes at all, but robots (Williamson, 2007, 87), etc. It might be that whoever possesses the ability to use 'vixen' in basic ways needs to assent to some sentences, but there is no fixed set of sentences to which anyone who possesses the ability to use 'vixen' in basic ways needs

[23] See also Williamson (2007, 97).

to assent. The ability to use ‘vixen’ in basic ways is, in this sense, “multiply realizable.” My view is that the same is not true in the case of formal mathematical systems. Let me explain.

What are the basic ways of using expressions of a formal language? To find this out, we should once again look at the normal context of use of formal languages. But, as we mentioned above when discussing Premise 1, the only use of formal languages is in the context of doing mathematics or metamathematics; formal languages are artificial devices used exclusively in these contexts. And in these contexts, the most basic ways of using a formal language are: doing basic formal derivations, and examining some of the most basic semantic properties of the formal system. There are no other basic uses of expressions of a formal language: One doesn’t apply words of a formal language to objects in the world, one doesn’t use formal languages in conversations, etc.

So, Understanding–Ability, when applied to formal languages, gets us that understanding expressions of a formal language requires being able to do basic formal derivations and basic semantic reasoning in that language. And, as I argued above in the context of showing that there are no expert counterexamples to formal analyticity (for Premise 2): Being able to do basic formal derivations or semantic reasoning in a formal language requires knowing the basic facts about the formal language, namely what are its basic axioms, rules of inference, and semantic properties.

So, understanding–ability links in the case of formal mathematical languages do imply understanding–assent links, in virtue of the fact that understanding expressions of the formal system requires being able to do some basic mathematics within the formal system, and that having the ability to do some basic mathematics within a formal system requires knowing at least the basic facts about the axioms, rules of inferences, and formal semantics of the formal system.

2.5 Looking ahead

In this paper, I introduced and defended a notion of analyticity that applies to formal languages by uncovering a flaw in Williamson's argument against analyticity when it is applied to formal languages. I argued that Williamson's argument, when applied to formal languages, in effect commits a fallacy of equivocation. For formal languages, the crucial notion of "assent" in Williamson's argument is ambiguous: On an internal reading, to assent to a sentence of a formal system is merely to recognize that it has a certain special status within the formal system (e.g. that it is true according to the formal semantics of the formal system, or that it is a derivation rule of the formal system). On an external reading, assent requires, in addition, that one takes the formal language in question to have a certain kind of representational adequacy. Philosophers like Graham Priest, who dissent from certain theorems or rules of inferences of formal systems, are not denying that those theorems or rules have a special status within the formal system; their contention is, instead, that the formal system is not adequate to represent certain kinds of correct reasoning. So they assent in the internal sense, and dissent only in the external sense.

Formal analyticity emerged as the notion of analyticity that involves the internal sense of 'assent': Understanding a formally analytic formula (or rule of inference) requires assenting to it being true or derivable (truth-preserving or a rule of inference) according to the rules of the formal system. I argued that there are formally analytic sentences of formal systems, such as disjunctive syllogism in FOL or Extensionality in ZFC: There won't be counterexamples—and certainly not expert counterexamples—who count as understanding these sentences without assenting to them in the internal sense. Finally, I argued that the reason why linguistic understanding for formal languages goes along with internal assent is that linguistic understanding for formal languages requires having the ability to do some basic mathematics within the formal languages, and that having this ability, in turn, requires knowing at least some basic facts about the rules and formal semantics of the formal languages.

Two questions are left for future investigation: (i) Can we find an analogue of formal analyticity in natural languages? and (ii) What are the epistemic properties of formally analytic sentences? I will conclude this paper by providing a roadmap for each question.

In Section 2.4, I proposed an intuitive assumption concerning necessary requirements on linguistic understanding, viz. Understanding–Ability, according to which understanding an expression requires having the ability to use it in basic ways. What counts as the "basic ways" to use an expression, I contended, is sensitive to the normal context of use for e. Formal analyticity emerged from a normal (and unique) context of use—formal mathematical practice—where having the ability to use e in basic ways requires having a fixed set of beliefs, Π_e. In order to look for an analogue to formal analyticity beyond formal languages, we should thus seek expressions whose normal contexts of use have basic requirements that are similarly "uniquely realizable." It is plausible, I think, that these can be found in informal areas of mathematics and in some other sciences.

That being said, there is a challenge for finding an analogue to formal analyticity in natural language. One key feature of formal languages is that they have only one context of use, namely formal mathematical practice. However, informal mathematics and sciences employ expressions that are also used in non-technical contexts, such as 'set' or 'electron'. For those expressions, we have to ask whether what counts as the ability to use the expression in basic ways depends on its use in the technical context or on its use in the non-technical context. If what counts as the ability to use e in basic ways depends on the use of e in the non-technical context, it will be hard to find sentences involving e that are similar enough to formally analytic sentences. But if what counts as the ability to use e in basic ways depends on its use in the technical context, or if e is ambiguous because it is used in such different contexts, then there may be enough similarity between some sentences involving e and formally analytic sentences that formal analyticity can model a notion of analyticity that applies to them.

Let me now turn to question (ii). In outline, formally analytic sentences of a given formal system are such that understanding them requires believing that they have some privileged status within that formal system. The concern here is that it is unclear why formally analytic sentences would thereby also have a privileged epistemic status. In my view, linguistic understanding at the level of formal languages does have epistemic import, although in a somewhat indirect manner. Let me explain.

A very plausible metasemantic account of formal expressions implies that axioms and rules of inference have a content-determining status.[24] It is quite plausible, in particular, to say that formal expressions of a formal mathematical language get their meanings in virtue of being associated with certain axioms and rules of inference by the relevant experts in the community: Once axioms and rules of inference get "fixed" in a community of mathematicians, they are thereby endowed with a content-determining status. The meaning of '$\in$', for instance, is at least partially determined by some of the basic axioms of ZFC. The meaning of '$\vee$', similarly, is at least partially determined by the rules and semantics of FOL. Notice that this leaves it open whether '$\in$' and 'is a member of' in natural language have the same meaning, or whether '$\vee$' and 'or' in natural language have the same meaning; the metasemantics of expressions of natural language may work quite differently from the metasemantics of expressions of formal languages.

Assume for a moment that this is the correct metasemantic account of expressions of formal languages, and that understanding alone can get us to believe (and even know) of certain propositions that they have a special status within the formal system. If one then knows the correct metasemantic account for these expressions—and thus knows

[24]This is of course how Carnap understood axioms (or "meaning postulates"), see for instance Carnap (1963, 1952a). See also, for instance, Boghossian (1996), Peacocke (2000), and Carey (2009).

that, in virtue of having this special status within the formal system, these propositions also have a content-determining status—then one can come to know something about the content of these expressions. For instance, if understanding '$\in$' is sufficient for knowing that Extensionality is an axiom of ZFC, and if one knows that axioms (at least the most basic axioms, say) of a formal system have a content-determining status, then we can come to know, on this basis, that sets are extensional.[25] Of course, an "external" question remains: We can ask whether the expressions of the formal language have the same semantic properties as those of certain corresponding expressions in natural language. We can ask, for instance, whether what we come to know about the content of '$\vee$' is also true about the content of 'or'. But we do at least come to know something about the content of the formal expressions.

Of course, the key remaining task is to find the correct metasemantic account for formal expressions. What would threaten the kind of epistemic import I sketched above is an account according to which none of the axioms or rules of inference of a formal system—not even the most basic ones—have a content-determining status. So for instance, in the case of set theory, it is a view according to which even expert users

[25]Perhaps one can also have justification or "entitlement" to believe axioms to be true even without knowing the correct metasemantic theory, just in virtue of understanding the expressions, and hence knowing that certain sentences have a privileged status within the formal systems. This is connected to Paul Boghossian's project: In a number of papers, Boghossian aimed to show that, if certain rules of inference have content-determining statuses, and if understanding the logical constants entails assenting to these rules of inference, then we can be justified or at least "blameless" in inferring according to these rules of inference (Boghossian, 1996, 2003a,b). Of course, the same strategy can't be used here because understanding expressions of a formal language, I argued, only requires "internal" assent, not something like "inferring according to the rules in ordinary reasoning."

of a formal language defer to, say, an optimal theory of sets according to which even Extensionality may be false about sets.[26] If such a view is correct, then knowing that Extensionality is an axiom of ZFC will be of no help for knowing that sets are extensional. There are a number of reasons to be skeptical of this alternative metasemantic account; for one, it would seem to go against the actual attitudes and intentions of expert users of formal languages—they most often don't intend to defer at least concerning the most basic axioms and rules of inference of some formal systems like FOL or ZFC.[27] In any case, it is here, at the level of metasemantics, that the question over the epistemic import of formal analyticity will, in my view, be decided.

[26]This kind of deference to an optimal theory is discussed in Rey (1998) and more recently in Smith (2015).

[27]Sheldon Smith makes a point similar to this in arguing against Rey's claim that Leibniz deferred to an optimal theory for the concept of the derivative (Smith, 2015, 15).

A picture held us captive. And we could not get outside it, for it lay in our language and language seemed to repeat it to us inexorably.

Ludwig Wittgenstein

3

Why Is the Universe of Sets Not a Set?

According to the iterative conception of sets, standardly formalized by Zermelo–Fraenkel set theory with Choice (henceforth 'ZFC'), there is no set of all sets. Following common usage, let 'the universe of sets' refer to the "totality" of sets in the iterative hierarchy of sets.[1] Two questions naturally arise: (i) What, if not a set, is the universe of sets? (henceforth 'the what-question') and (ii) Why is the universe of sets not a set? (henceforth 'the why-question'). The why-question will be the guiding question of this paper, and investigating it will require us to look at the what-question. Note that there are equivalent what- and why-questions about other "totalities" that are not members of any sets, such as the "totality" of all ordinals; although I won't directly address those questions in this paper, the discussions that follow could be generalized to apply to them as well.

Here is a simple-minded response to the why-question:

Minimal Explanation The universe of sets is not a set because the assumption that it is contradicts some axioms of ZFC.[2]

There are a number of different ways to derive a contradiction from the assumption that the universe of sets is a set—e.g. via the Foundation and Pairing axioms, the Separation axiom, the Burali-Forti argument, Cantor's Theorem—and each one generates a more specific instance of the minimal explanation. To have a clear example in mind, let me expand the instance of the minimal explanation generated by the proof from Founda-

[1] I intend the expression 'the universe of sets' to be neutral with respect to the nature of the "totality" of all sets in the iterative hierarchy. As we will see in this paper, the universe of sets may be a proper class, a plurality, an incomplete totality, a potential hierarchy (in which case there does not actually exist a plurality of all sets), etc.

[2] The discussion can be generalized to other why-questions by formulating equivalent forms of the minimal explanation. So for instance, the equivalent "minimal explanation" for the why-question about the "totality" of all ordinals would be: "The totality of all ordinals is not a set because the assumption that it is contradicts some axioms of ZFC."

tion and Pairing (henceforth 'the Minimal Explanation$_{\text{F, P}}$') by spelling out how the contradiction is derived: "The universe of sets, V, is not a set because the assumption that it is contradicts Foundation and Pairing, as follows. If you assume that V is a set, then $\{V\}$ is also a set (by Pairing). Thus $\{V\}$ must have a member that is disjoint from $\{V\}$ (by Foundation). The only member of $\{V\}$ is V, hence $V \cap \{V\} = \emptyset$. But $V \in V$ (by definition of V), so $V \cap \{V\} \neq \emptyset$. Contradiction."

Many philosophers seem to think that the minimal explanation is not a good answer to the why-question. In this paper, I first precisify the core complaint against the minimal explanation Section 3.1), and then argue against the two main alternative answers to the why-question: the size-explanation (Section 3.2) and the potentialist explanation (Section 3.3). The problems with the alternative answers give us reason to reconsider the minimal explanation as an answer to the why-question. I conclude the paper by outlining a close alternative to the minimal explanation, the conception-based explanation, that avoids the core complaint against the minimal explanation.

3.1 The Core Complaint Against the Minimal Explanation

Philosophers frequently repeat the why-question right after they have explained some instance of the minimal explanation. Here is Keith Simmons doing just that:

> [W]ith the combinatorial/iterative conception in mind, why can't we "collect together" or "lasso" all the sets in the ZF hierarchy, and form the collection of them all? (Simmons, 2000, 111)

More recently, right after stating the instance of the minimal explanation with Separation, James Studd asks:

> What is it about the world that allows some sets to form a set, whilst prohibiting others from doing the same? (Studd, 2013, 699, emphasis removed).

It is natural to take such philosophers to think that the minimal explanation doesn't explain why the universe of sets is not a set. Some—like Michael Dummett here concerning the instance of the minimal explanation with Cantor's Theorem—make this thought more explicit:

> A mere prohibition leaves the matter a mystery. …And merely to say, "If you persist in talking about the number of all cardinal numbers, you will run into contradiction," is to wield the big stick, but not to offer an explanation. (Dummett, 1991, 315–316)[3]

However, it is hard to find a clear statement of what exactly is wrong with the minimal explanation as an answer to the why-question. The following two passages contain what I take to come closest to a precise complaint against the minimal explanation.

Right before stating his own potentialist answer to the why-question, Stephen Yablo states:

> This [Minimal Explanation$_{\mathrm{F,\ P}}$] brings what the Russell set is by nature into conflict with a basic fact about sets, viz. well-foundedness. But one may question whether the fact is basic enough. To say that …[the universal set] cannot exist because it would be ill-founded seems to get things the wrong way around. It is because sets like …[the universal set] are independently problematic that we are drawn to a requirement that keeps those sets out. (Yablo, 2004, 149)

[3]See also for instance Priest (1979, 219f.).

One way to understand Yablo here is as follows: There is some fact about the universe of sets that explains both why it is not a set and why we adopt Foundation as an axiom (i.e. why "we are drawn to a requirement that keeps those sets out" (ibid.)). So, even though we can derive that the universal set doesn't exist from Foundation, we have to provide that "deeper" fact in explaining why the universal set doesn't exist. That fact, as Yablo later goes on to propose, concerns the "potential" nature of the universe of sets.[4]

In the same vein, James Studd, who also gives a potentialist answer to the why-question, suggests that the answer to the why-question ought to appeal to something about *sets*, and not merely to what can be derived within certain theories:

> The derivation of Russell's paradox in Naïve Set Theory demonstrates the logical falsity of the instance of the Naïve Comprehension schema .... This provides—I am happy to grant—as good an explanation as we should expect for why this *theory* is inconsistent. However, the question of real interest is not why this instance of Naïve 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. (Studd, 2013, 700)

I thus propose we understand the core complaint against the minimal explanation as follows. The minimal explanation is a perfectly good answer to the following question: "Why is the universe of sets not as set according to ZFC?" But the minimal explanation fails to provide the deeper reason why these axioms that prohibit the existence of the universal set are there in the first place. So the minimal explanation is not a deep

[4]See for instance Yablo (2004, 152–155).

enough explanation of why the universe of sets is not a set; it is not a good answer to the why-question.

Understanding the core complaint this way also helps make sense of the alternative answers to the why-question. Indeed, most philosophers propose to answer the why-question by appealing to some "deep" fact about the nature of the universe of sets—one that may also provide some independent motivation or explanation for the axioms of set theory that prohibit the existence of the universal set. Their answers naturally go with either of two theories concerning the deep nature of the universe of sets: actualism combined with an appeal to the Limitation of Size Principle, and potentialism. In what follows, I will argue in turn against each of these answers.

3.2 Actualism and Size

For lack of standard terminology, let me use 'actualism' as the name of the general thesis that the universe of sets is some sort of "completed totality." Actualism is thus an answer to my what-question. Proponents of actualism include George Boolos, John Burgess, David Lewis, Donald A. Martin, Gabriel Uzquiano, and arguably early set theorists like Abraham Fraenkel, Kurt Gödel, and John von Neumann. Some actualists take a "completed totality" to be a proper class: a well-founded extensional object that is not a set.[5] Others take it to be a plurality that cannot be "singularized" into either

[5] See for instance Welch (ms), Lewis (1991), and Mayberry (1986).

a set or a proper class.[6]

Many actualists accept some version of the Limitation of Size Principle, if only because it is derivable in most set theories with proper classes or pluralities:[7]

Limitation of Size Principle A collection, C, is not a set if and only if it is of the same size as V (the totality of all sets).

A number of actualists appeal to the Limitation of Size Principle to justify some of the axioms of ZFC, by stating that an axiom that asserts the existence of sets with a certain property is true if the relevant sets are "small" enough.[8] Similarly, such actualists use the Limitation of Size Principle to provide an explanation for why the universe of sets is not a set: The universe of sets is not a set because it is "too large" to be a set. Let me call this 'the size-explanation', and actualists who endorse the size-explanation 'size-actualists'. Many mathematicians are size-actualists, including people like von Neumann, Cantor, and Mirimanoff.[9] Even though it is hard to find philosophers who argue for the size-explanation at length, people like John Burgess, Philip Welch, and arguably David Lewis at least commit themselves to it.[10]

---

[6] See for instance Boolos (1984), Burgess (2004), and Uzquiano (2003).

[7] The derivation requires some version of the axiom of Global Choice. See for instance Linnebo (2010, 162) for the derivation of a limitation of size principle from plural set theory.

[8] See for instance Fraenkel et al. (1973). For a thorough study of the history of the Limitation of Size Principle from Cantor to von Neumann and beyond, see Hallett (1984).

[9] Size-actualism is defended in Fraenkel et al. (1973) and by Georg Cantor, John Von Neumann, Dmitry Mirimanoff and other Limitation of Size-Theorists, as explained in Hallett (1984). More recently, Welch (ms, 9) assumes some version of the size-explanation.

[10] See for instance Burgess (2004) and Lewis (1991), provided as examples of size-actualists by Linnebo (Linnebo, 2010, 151, fn. 9).

Let me put my cards on the table: I think that actualism is the right view concerning the nature of the universe of sets (and so it is the right answer to the what-question), but I don't think that the size of the universe of sets has anything to do with the reason why there is no set of all sets. So what I will argue here is that the size-explanation is not a good answer to the why-question.

3.2.1 The Size-Explanation

Let us call the size of V, 'T'. T is the threshold size at which collections are too big to form sets. Some philosophers have already objected to the size-explanation on the grounds that T is an *arbitrary* threshold, especially in light of the practice of set theorists to accept larger and larger cardinals. See for instance Øystein Linnebo:

> The main challenge [for the actualist] will be to motivate and defend [T] …. Why should this particular cardinality mark the threshold? Why not some other cardinality?
>
> …Wherever it has been possible to go on to define larger sets, set theorists have in fact done so. So it remains arbitrary that there should be no sets of this cardinality or some even larger one. (Linnebo, 2010, 152–153)[11]

While I reject the size-explanation, I don't think this is the right objection to make against it; it mischaracterizes the way actualists think (and should think) of the threshold T.

To see this, the first thing to note is that for set theorists, κ is a *cardinal* if and only if κ is an ordinal, and for every $\eta < \kappa$, there is no bijection between η and κ (i.e. $\eta \not\approx \kappa$).

[11] See also Linnebo (2013, 206): "To disallow such a set [of all sets] would be to truncate the iterative hierarchy at an arbitrary level."

The *cardinality* of a given set A, written '$|A|$', is in turn defined as the least ordinal κ such that $A \approx \kappa$.[12] Since all ordinals are sets, crucially, all cardinals and cardinalities are sets.

The actualist accepts these definitions. She then agrees with Linnebo that there are sets of larger and larger cardinalities, as is borne out by the practice of set theory. The important point comes now: The actualist doesn't think of T as a cardinality in this sense. The actualists can't both maintain that there are sets of larger and larger cardinalities, and that there is a threshold cardinality above which there are no sets of larger cardinality: That would not be arbitrary, it would just be inconsistent.

The actualist needs to think of the "size" T differently. For instance, she may define 'same size' by stating that A has the *same size* as B if and only if $A \approx B$, and at least partially define 'size' by stating that (a) if A has a cardinality $|A|$, then the *size* of A is $|A|$, and (b) if A doesn't have a cardinality and has the same size as V, then the *size* of A is T. So for the actualist, to say that something X has size T just is to say that $X \approx V$; the actualist can't specify T in any other way. Now, the "location" of T is certainly not arbitrary, because T is the first size at which the assumption that A has size T implies that A is not a set. Moreover, the existence of larger and larger cardinalities doesn't have any bearing on the "location" of T. It is true that for any cardinality κ, there are sets A of cardinality greater than κ. It is also true that every

[12] This definition requires the Axiom of Choice. Without Choice, one can use Dana Scott's trick to define $|A| = \{X \mid X \text{ is of minimal rank s.t. } A \approx X\}$. Since the collection of sets of a given rank is always a set, cardinalities are always sets according to this definition.

cardinality is a size. But it surely doesn't follow from these two facts that for any *size* S, there are *sets* of size greater than S.

Notice that at this point, one could raise a variant arbitrariness complaint: "It would be arbitrary to maintain that for any cardinal κ, there are sets A of cardinality greater than or equal to κ, but that for some size T, there are no sets A of size greater than or equal to T." The actualist should respond that there simply are no sets A that have size greater than or equal to T, on pain of contradiction. She may appease her opponents by accepting that there are collections of size greater than T (or even so-called "Super-Classes," collections of size greater than classes that are neither sets nor classes).[13] In other words, she may accept that just as there are larger and larger cardinalities, there are larger and larger sizes. In any case, though, her crucial point is that not all sizes are sizes of *sets*, on pain of contradiction, and hence that, yet again, nothing about T has been shown to be arbitrary.

So the threshold is not arbitrary, because it is not specifiable (for instance, as a particular cardinality). And I think the real problem with the size-explanation has to do with precisely this, with T not being a *specifiable* threshold. To see this, let us first consider a good explanation that appeals to a threshold size and that looks, at least on the face of it, quite similar to the size-explanation. If you scaled me up (that is, if you increased my height while keeping my shape the same), there is a height H above which my body would collapse under its own weight. Assume I in fact just collapsed.

[13]Such entities are discussed for instance in Lévy (1976).

Now consider the following question: "Why did I collapse under my own weight?" The answer to this question will appeal to the "square-cube law," namely that if my height is scaled-up by multiplier m, then my surface area is scaled up by m^2 and my volume by m^3. The reason why I collapsed when I reached height H is that my bone strength is (roughly) proportional to the surface area of the cross section of my bones, but weight is a function of volume, hence after H, my bones won't be able to support my increased weight. That's a good explanation in terms of a threshold size.

The size-explanation is not like the explanation of my collapse. In the case of the size-explanation, we don't have any independent way of specifying the size of V. One thing we can say about T is that it is the size of all collections that aren't sets on pain of contradiction. But if that's what we say about T, then the size-explanation is obviously circular. It would be like saying that I collapsed under my own weight because I had height, H, and the most we could say about H is that it is the height at which I collapse under my own weight.

What if we said that T is just the size of V? There is this size, the size of V, and nothing that has the size of V is a set. This, too, should strike us as a very poor explanation. It would be like saying that I collapsed under my own weight because I had height H, and the most we could say about H is that it was the height that I had. In that case, being told that I had size H doesn't make it intelligible why I collapsed under my own weight; there is no way to "connect" the size H to the event of my collapsing under my own weight by appealing to certain background laws. The explanation for why

I collapsed under my own weight after height H made it intelligible how H is connected, via background laws of physics, to my collapse. But the size-explanation doesn't make it intelligible how the size of V has anything to do with its not being able to form a set. The size-explanation is just not a good explanation.[14]

3.3 Potentialism

For lack of standard terminology, let me use 'potentialism' as the name of the general thesis that the universe of sets is an "incomplete" or "potential" totality. More exactly, for the potentialist, necessarily, no matter how many stages of the hierarchy have been formed, it is always possible that there be a further stage containing sets whose members do not form sets in any of the preceding stages. Potentialism is thus another answer to my what-question. Its proponents include Ignacio Jané, Geoffrey Hellman, Øystein Linnebo, Charles Parsons, Augustín Rayo, James Studd, William Tait, Stephen Yablo, and arguably Kit Fine and Ernst Zermelo.[15]

Potentialists usually see their account as providing a motivation for the ordinary axioms of set theory. We saw this in Yablo's quote in Section 3.1. Similarly, Studd says

[14] In a recent paper, Christopher Menzel argues that the iterative conception of sets is intuitively consistent with the existence of a proper-class sized set of ur-elements, and proposes a modification to Replacement and Powerset to accommodate these "wide" sets (Menzel, 2014). Menzel (2014) may thus be seen as providing an alternative argument for the insignificance of size in explaining why the universe of sets is not a set: the universe of sets is still not a set according to Menzel's modified formalization of the iterative conception, but this has nothing to do with the "size" of the universe of sets.

[15] See for instance Jané (1995), Hellman (1989) and Zermelo (1930). The other potentialists are cited below.

that his aim is to "recover the modal analogue of ZF" from the potentialist principles, just like Linnebo (Studd, 2013, 5). Here is how Linnebo puts the point:

> [W]hen we confront difficult foundational and conceptual questions concerning set theory, the finer resolution provided by the modal approach can be very valuable. In particular, we will see that the modal approach makes available a very natural motivation for the axioms of ZF set theory. (Linnebo, 2013, 206)[16]

So potentialists purport to be providing a "deep" account of the nature of the universe of sets in the sense of Section 3.1.

Moreover, almost all potentialists take potentialism to provide an explanation for why the universe of sets is not a set: The universe of sets is not a set because the universe of sets is a "potential" totality. Let us call this the 'potentialist explanation'. See for example Parsons, who makes clear that potentialism provides an answer to the why-question; he does this in the context of evaluating Georg Cantor's distinction between "definite multiplicities" and "inconsistent multiplicities" (Cantor, 1899):

> I suggest interpreting Cantor by means of a modal language with quantifiers, where within a modal operator a quantifier always ranges over a set .... Then it is not possible that all elements of, say, Russell's class exist, although for any element, it is possible that *it* exists. (Parsons, 1977, 515)

In the same vein, Linnebo makes clear that he endorses what he takes to be Cantor's answer to the why-question:

[16]See also Studd (2013, 698f.) where Studd says of Boolos' stage theory (Boolos, 1971) that it provides a good motivation for the axioms of set theory but is not able to preserve the claim that every sets can form a set nor answer the why-question. Studd then proposes modal set theory as an alternative to stage theory.

> On [the potentialist] …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 …a reason to disallow the disputed set formation. (Linnebo, 2013, 206)

> [One] attraction of the potentialist conception emerges in connection with the hard question of the conditions under which some objects are eligible to form a set. …[The potentialist's] thought is that there is an intrinsic difference between multiplicities that form sets and multiplicities that do not, and that this intrinsic difference *explains* why some but not all multiplicities are eligible for set formation. (Linnebo, 2013, 206–207)[17]

What I want to do here in Section 3.3 is to criticize the potentialist explanation. First, let me explain the potentialist thesis in a little more detail.

3.3.1 The Potential Hierarchy

When we explain the iterative conception of sets, we usually speak as though sets are formed in time. At the beginning we form no sets: $V_0 = \emptyset$. At the next stage we form the set of all sets formed so far: $V_1 = \mathcal{P}(V_0) = \{\emptyset\}$. Continuing in this way we obtain $V_2 = \mathcal{P}(\mathcal{P}(V_0))$, $V_3 = \mathcal{P}(\mathcal{P}(\mathcal{P}(V_0)))$, ... After all the finite stages, we form the set of everything that came before, namely $V_\omega = \bigcup_{n<\omega} V_n$. We can then start taking powersets again, until the next limit stage, $\omega + \omega$, at which we form the set of everything that came before, namely $V_{\omega+\omega} = \bigcup_{\alpha<\omega+\omega} V_\alpha$, ... and so on, thereby forming the cumulative hierarchy of sets (Figure 3.1).

[17] In his recent defense of potentialism, Studd similarly takes potentialism to provide an answer to the why-question; he states that potentialism is supposed to answer the "difficult question" I cited above on page 100.

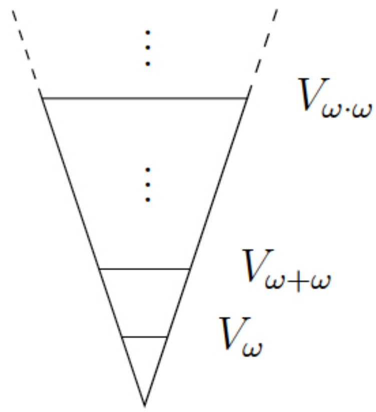

Figure 3.1: The cumulative hierarchy of sets.

Potentialists take this informal explanation, "replace the language of time and activity with the more bloodless language of potentiality and actuality" (Parsons, 1977, 526), and then formalize it; their aim is to regiment a way of talking about sets that includes claims about possibility, in particular the claim that any sets (any zero or more sets) *can* form sets.[18] To this end, they propose formal systems of set theory supplemented by modal quantifiers '$\Diamond\exists$' and '$\Box\forall$', and either plural logic (Linnebo), second-order logic (Parsons), or additional "backwards-looking" modal operators (Studd). Neither formalization adds any new results about sets: a formula φ is derivable in ZF if and only if the formula $\varphi^\Diamond$ that results from replacing each quantifier in φ with its corresponding modal quantifier can be proved in modal set theory.[19] According to potentialists, this shows that "in

[18] Studd calls this 'the Maximality thesis' (Studd, 2013, 699), Linnebo calls it '(C)' (Linnebo, 2013, 219), and Parsons introduces it in Parsons (1977, 527).

[19] See for instance Linnebo (2013, 214) and Studd (2013, 710).

the context of [their] modal set theory the composite expressions $\Box\forall$ and $\Diamond\exists$ behave logically just like ordinary quantifiers," and is a reason why "the implicit modalities [they] have postulated in the set-theoretic quantifiers do not surface in ordinary set-theoretic practice" (Linnebo, 2010, 164).

In outline, the potential hierarchy is a structure of possible worlds with an accessibility relation that corresponds to S4.2 (Figure 3.2[20]).[21]

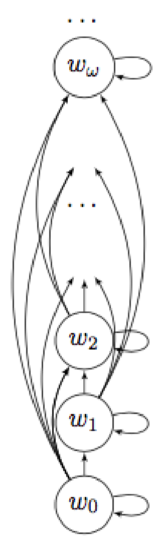

Figure 3.2: The potential hierarchy of sets.

Worlds are stages of the set forming process, the domain of each world consists of the sets that have been formed thus far, and the domain of each world accessible from a

[20]Note that, although Figure 2 suggests otherwise, the accessibility relation in S4.2 is not a linear relation; Figure 2 is merely intended to provide a rough visual idea of the potential hierarchy of sets.

[21]In particular, Studd's modalization is different from Linnebo's in that it has two basic necessity operators '$\Box_<$' and '$\Box_>$', and derived ones '$\Box$', '$\Box_\leq$' and '$\Box_\geq$', corresponding respectively to S5 and S4.3 (for the last two). Here I set aside Studd's formalization without loss of generality for my arguments in Sections 3.3.2–3.3.3.

given world w_α is a superset of the domain of w_α (Linnebo, 2013, 208). From each world w_α, there are sets that don't exist at w_α but that are possible relative to w_α, i.e. sets that exist in the domain of a world w_β which is accessible from w_α. The overall structure of the potential hierarchy is perfectly isomorphic to that of the cumulative V_α-hierarchy. Moreover, modal set theory and non-modal set theory are "mutually interpretable," in that modalized versions $\mathrm{ZF}^\Diamond$ of the ZF axioms are true of the potential hierarchy of sets,[22] and modal set theory can be interpreted in non-modal set theory (hence is also consistent relative to it).[23] One way to think about the difference between ordinary set theorists and potentialists is that the latter have "powerful instruments for studying the same subject matter under a finer resolution" (Linnebo, 2013, 206); that is, they "uncover" accessibility relations "between" stages V_α.

Let me now turn to the potentialist's answer to the why-question. In the potential hierarchy, there is no universal set—there is no world w_V at which all sets exist; quantification over sets is always quantification over the domain of a world, but from each world it is always possible to go on to form more sets, and hence there is no "totality" of all sets in any one world. This is the potentialist's answer to the why-question (the potentialist explanation).

Problems for the potentialist explanation arise once one carefully examines the role

[22]See for instance Linnebo (2013, 220ff.), Parsons (1983, 318ff.), and Studd (2013, 712ff.).

[23]In particular the claim (C) that any sets (any zero or more sets) can form a set is mapped on to the claim that for any stage α, and any subset of V_α, there is a later stage β such that all later stages γ contain a set containing all and only those things in the subset of V_α (Linnebo, 2013, 224).

and nature of the potentialists' notion of modality, which is key in their explanation. All potentialists agree that their modality is not metaphysical modality: since pure sets and mathematical objects exist necessarily, there is nothing metaphysically potential about the hierarchy of sets.[24] Instead, they either take the modality as a primitive notion idiosyncratic to mathematics, or interpret it to track set theorists's practice to accept sets of larger and larger cardinalities.[25] These two options generate two slightly different versions of potentialism. In what follows, I evaluate the potentialist explanation under each option.

3.3.2 Primitive and Idiosyncratic Modality

The first option is to take the potentialist's modality as "mathematical modality," a primitive notion idiosyncratic to mathematics (Parsons, 1983, 327). Linnebo similarly proposes:

> Strictly speaking, all we need to assume about the above notion of modality is that it is suited to explicating the iterative conception of sets. The modality must thus be one on which the existence of a set is potential relative to the existence of its elements (in the sense that, when some things

[24]See for instance Linnebo (2010, 158), Studd (2013, 706), Linnebo (2013, 226), Fine (2006, 31), and Parsons (1983, 328).

[25]Examples of the former include Linnebo (2013) and Parsons (1977). Linnebo considers the latter option in Linnebo (2010, 159). Some potentialists simply propose to forgo specifying the modality: "A full-fledged explanation of the modal notions will have to await another occasion" (Linnebo, 2013, 207). See also Studd: "There is a great deal more to be said about each of these views, but it would take us too far afield to say it here. Rather, safe in the knowledge that taking the tense more seriously than usual need not commit us to taking it literally, I shall continue to elaborate on this view in general, leaving it open (within the bounds of LST) how the modality is to be interpreted" (Studd, 2013, 707).

> exist, it is possible for there to exist a set with precisely these things as elements.) All other details are optional. (Linnebo, 2010, 158)

In this subsection, I will argue that the potentialist explanation for the why-question on this first option for interpreting the potentialist's modality is redundant. I will do so in two steps.

The first step is to notice that the potentialist's modalities come at a cost: since ordinary set theorists don't use modal quantifiers, potentialists buy the distance between the axioms (and the why-question) and what explains them at the cost of having to reinterpret the practice of set theorists. Potentialists need to establish a correspondence between the practice of set theory and the "worlds" of the potential hierarchy, and, in particular, specify the sets that are "actual" according to their modality. Linnebo considers this difficulty and proposes two options for the potentialist. The first option is to interpret set theorists to be working "at the world" which contains sets "whose existence follows from our strongest, well-established set theory" (Linnebo, 2010, 159, fn. 21). I will examine this option in Section 3.3.3, as it is embodied in a slightly different version of potentialism. The second option is not to specify any "actual" world but instead to assign set theorists an "external" perspective on the potential hierarchy of sets:

> [S]et theorists generally do not regard themselves as located at some particular stage of the process of forming sets but rather take an external view on the entire process. It therefore would be wrong to assign ourselves any particular stage of the process. (Linnebo, 2010, 159)

I will focus on this second option for the remainder of this subsection.

Assume set theorists in their practice take an "external" perspective on the potential hierarchy. Note that potentialists, too, allow themselves such a perspective: in order to prove Infinity$^\Diamond$, for instance, Linnebo and Studd advocate a "reflection principle," $\varphi^\Diamond \to \Diamond\varphi$, which is understood as the claim that

> [T]he truth of a claim in 'the model' provided by the potential hierarchy of sets ensures that the claim is possible. For a claim φ to be true in this 'model' is for φ to be true when all its quantifiers are understood as ranging over all possible sets, including ones not yet formed. But for φ to be true when understood in this way is simply for its potentialist translation $\varphi^\Diamond$ to be true. (Linnebo, 2013, 222)

But if we are allowed to take such an "external" perspective on the whole potential hierarchy, and even understand our quantifiers as ranging over all possible sets, then the why-question simply resurfaces. Remember that potentialists are quick to ask the why-question concerning the cumulative hierarchy: "Why can't we 'collect together' or 'lasso' all the sets in the cumulative hierarchy and form a set of all sets?" Presumably, potentialists imagine drawing a circle around Figure 3.1 and want to know why that doesn't represent a set. But why couldn't we ask the same question with respect to the whole potential hierarchy of sets seen from this "external" perspective? Imagine drawing a circle around Figure 3.2. Can't we now ask: "Why doesn't that represent a (possibly existing) set?"

To answer the why-question for the cumulative hierarchy, the supporter of the minimal explanation can reply that, strictly speaking, set theory happens "inside" the universe V, that although we can "look" at V from the outside this doesn't mean that it is a set, and that in fact V is not a set on pain of inconsistency with set theory. The potentialist is not satisfied with this answer. But what, besides the claim that modal set theory precludes the existence of a world at which all sets exist, can the potentialist offer as an answer to the why-question for the potential hierarchy? She will similarly point out that sets strictly exist "inside" worlds, that although we can (and do) look at the whole hierarchy from the "outside" this doesn't mean that we can form its set, and that modal set theory is designed so that assuming that a universal set is possible entails a contradiction.[26] So the potentialist will have to provide us with a version of the minimal explanation from within modal set theory, instead of providing the minimal explanation from within ZFC.

This is not yet a devastating problem for the potentialist. Indeed, potentialists may reply that modal set theory captures some deeper truths about sets than does ZFC, and hence claim that the minimal explanation within modal set theory is more satisfactory than the minimal explanation within ZFC. But this is where the potentialist's primitive and idiosyncratic notion of modality causes trouble. The second step of the argument is thus to note, as we did above in Section 3.3.1, that the potential and iterative hierarchies are perfectly isomorphic, and modal and non-modal set theories are mutually interpretable.[27] This means we cannot get any grip on the potentialist's modality by

[26] See Linnebo (2013, 222f.) for one example of this kind of "design."

[27] One further result deserves mention here besides the one in Section 3.3.1. If φ is a formula all of whose quantifiers are modalized, then modal set theory proves that φ, $\Box\varphi$ and $\Diamond\varphi$ are

merely considering the set of true sentences containing '□' and '◊'. 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 as stages V_α defined in ZFC? What exactly is added by the '□' and '◊' 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: to say that the universe of sets is not a set because it is "potential" in that at any stage, we "can" form more sets in this unspecified and idiosyncratic sense of "can" is not far from giving a dormitive virtue explanation, or saying nothing at all. In other words, simply having unexplained '□' and '◊' in front of the quantifiers in the minimal explanation doesn't make the potentialist explanation any deeper or more informative than the minimal explanation. The potentialist explanation with this unexplained, primitive and idiosyncratic notion of modality doesn't provide any deeper insight on the why-question than the minimal explanation.

We have just seen that in order to answer the why-question, potentialists commend a costly detour through modal set theory. On the way of incurring the cost just considered, their explanations ultimately rely on a minimal explanation and add nothing to it. I will argue the same for the second way of incurring the cost in Section 3.3.3.

equivalent (see e.g. Studd (2013, 709) and Linnebo (2013, 213)), meaning that it doesn't matter at which world a full modalized formula is evaluated, and which is "another reason why the implicit modalities that I have postulated in the set-theoretic quantifiers do not surface in ordinary set-theoretic practice" (Linnebo, 2010, 164).

3.3.3 Interpreted Modality

The second option is to interpret the potentialist's modality as somehow tracking the practice of set theorists to accept the existence of larger and larger cardinals. As we saw in Section 3.3.2, Linnebo considers this option in an attempt to re-interpret the practice of set theorists:

> As science progresses, we formulate set theories that characterize larger and larger initial segments of the universe of sets. At any one time, precisely those sets are actual whose existence follows from our strongest well-established theory. (Linnebo, 2010, 159, fn. 21)[28]

The version of potentialism associated with this interpretation of the modality is subtly different from the one we considered previously. My criticisms in this subsection will only concern the main idea behind this version of potentialism, so I will keep my exposition at a high level of generality; more detailed accounts can be found in Koellner (msa), Parsons (1974), Rayo & Linnebo (2012) and Tait (2005).

In outline, potentialists of this variety claim that sets exist if and only if they are specified by a theory (call this claim 'Specifiability'). The motivation for Specifiability may be a more "constructivist" or "postulationist" metaphysical outlook. Some potentialists put constraints on the kinds of theories that can specify sets in Specifiability. Tait, for instance, mentions the requirement of categoricity:

> [W]e may speak about the existence of this or that object in mathematics only when we have specified a consistent and categorical theory in which we speak of such objects. (Tait, 2005, 141)

[28]See also Linnebo (2013, 207f.).

Others simply let the theories in question be the ones accepted by set theorists.

The second main claim these potentialists endorse is that theories can always be expanded (call this 'Expandability'). Then, Specifiability and Expandability together provide the potentialist with an answer to the why-question: Since sets exist only if they are specified by theories, and since theories can always be expanded, there is no one completed totality of "all sets" which can form a set. In other words, just as it makes no sense to "collect together" all theories, it similarly makes no sense to "collect" together the universe of sets. Here is a quote I take to be voicing this idea:

> [The potentialist] view makes it impossible to draw a clean separation between the question of how one might extend one's expressive resources and the question of how many sets exist. By increasing one's expressive resources in the right sort of way, one is led to recognize additional ontology. So insofar as one believes that the process of extending one's expressive resources is essentially open-ended, one should also think that the hierarchy of sets is essentially open-ended—and therefore that there is no definite fact of the matter about what sets there are. (Rayo & Linnebo, 2012, 292)

Once again, there are various ways to understand Expandability. Some philosophers, like Linnebo in the quote above, take theories to be "expandable" in that set theorists define and accept larger and larger cardinals (e.g. Inaccessible, Mahlo, Measurable, Strong, Woodin, Supercompact, etc.)[29] One problem with this understanding is that it risks making the expandability of theories—and thereby also the non-existence of a set of all sets—contingent on the particular choices and activities of set theorists. Would there be a universal set if set theorists stopped accepting the existence of new cardinals?

[29]See Koellner (2011) for more on the Large Cardinal Hierarchy.

Surely not, since a universal set would still be inconsistent.[30] A more plausible alternative is to say that theories are "expandable" in that for any recursively enumerable theory T_n we can define $T_{n+1} = T_n +$ "there is an inaccessible κ such that $V_\kappa \models T_n$," or that any theory of the form ZFC + $\exists\varphi$-cardinals can be "expanded" to ZFC + "There is an inaccessible cardinal κ such that V_κ satisfies that there is a φ-cardinal."[31]

Now that we have a clearer picture of the potentialist explanation under this second option for interpreting the potentialist's modality, I want to raise a problem for it that applies on either way of precisifying Specifiability and Expandability. The problem is that Expandability and Specifiability are compatible with the existence of a universal set, hence they cannot explain why there is no set of all sets. Indeed, there are set theories with universal sets (i.e. where it is an axiom that there is a set of all sets) for which the same expandability phenomenon applies: one can define larger and larger cardinals within such theories.[32] To take a specific example, NFU^+ (that is, Jensen's version of Quine's NF modified to accommodate urelements, called 'NFU', supplemented by Infinity and Choice) can be expanded as far as one is willing to expand ZFC (Holmes,

[30] There are many other problems facing a radical constructivist view on which the existence of sets depends on whether set theorists define them or think about them. For one, it is widely accepted that such a radical constructivist approach will only sanction a much weaker and non-classical set theory. But actualists and potentialists don't want to be revisionists—as we have seen, they want to "recover" most of the axioms of ZF. So I set aside this kind of revisionist option in this paper. For a recent discussion and summary of the arguments against constructivism, see for instance Incurvati (2012).

[31] This way of making sense of expandability fits with Tait (2005) and Parsons (1974).

[32] For exposition of set-theories with universal sets, see Forster (1992), Holmes (1998) and Holmes (2001).

2001).[33] Nothing about those theories conflicts with Specifiability. Thus, since Expandability and Specifiability are compatible with the existence of a universal set, they cannot explain the absence of a universal set.

Let me consider a response on behalf of the potentialist. The potentialist may state: "We are not concerned with set theories like NFU^+. We are only interested in set theories that formalize the iterative conception of sets, and hence that have Foundation. NFU^+ doesn't have Foundation. Our claim is that *within* well-founded theories, it is their expandability that explains why there is no set of all sets."[34] The potentialist can state the same response with Separation instead of Foundation, for instance. Without loss of generality, I will only focus on the response with Foundation. I now want to argue that this response is unsuccessful, as it renders the potentialist explanation either once again redundant or needlessly costly.

As we know, if we restrict our attention to theories that formalize the iterative conception—and, in particular, to theories that have Foundation—then we have Minimal Explanation$_{\mathrm{F,\ P}}$; we can explain why the universe of sets is not a set with a one-line mathematical argument. Instead, the potentialist wants to add a detour through Expandability. But this is an unnecessary detour, given that Expandability by itself is

[33]NFU is consistent (relative to PA) and NFU^+ is consistent relative to Zermelo set theory with only Δ_0-Comprehension (Jensen, 1969). NFU^+ + the Axiom of Cantorian Sets (which says that all Cantorian sets are strongly Cantorian) proves the existence of inaccessible cardinals, and n-Mahlo cardinals for each n. A *Cantorian set* is a set A such that $|A| = |\mathcal{P}_1(A)|$, where $\mathcal{P}_1(A)$ is the set of all one-element subsets of A, and a set A is *strongly Cantorian* if the class map $(x \mapsto \{x\}) \upharpoonright A$ is a set.

[34]Potentialists usually make clear that their focus is on the iterative concept of set, so this kind of response is not implausible for a potentialist. See for instance Linnebo (2010, 144).

compatible with the existence of a universal set; Expandability is an idle wheel in the explanation of why there is no set of all sets. The core problem with the universal set is not that set theories such as ZFC can be expanded: the core problem is that a universal set would be inconsistent with the assumptions of a theory that formalizes the iterative conception of sets, such as ZFC.

One might object to this argument by denying something like the following principle, which seems to be implicit in it: If a fact p is compatible with q not obtaining, then p cannot explain q.[35] However, a related worry faces the potentialist explanation whatever stance one takes towards this principle. At this stage of the dialectic, potentialists face two alternative explanations: the first one (i) appeals to Expandability, Specifiability, Foundation and Pairing, the second one (ii) (Minimal Explanation$_{\text{F, P}}$) appeals to Foundation and Pairing. If the potentialist insists on providing explanation (i), then she incurs costs both because (i) is much more complicated than (ii), but also because (i) contains an extra assumption, Specifiability, that needs to be defended. Indeed, why think that sets are specifiable totalities in the first place? This assumption is not even part of the iterative conception of sets. The potentialist here buys the explanatory distance at the cost of having to defend a substantial metaphysical assumption about the nature of sets.

[35] In the literature on grounding, people call this principle 'the entailment principle' (Rosen, 2010). Eli Chudnoff (ms) and Jonathan Dancy (2004) are among philosophers who deny this principle.

3.4 The Conception-Based Explanation

In Sections 3.2–3.3, I argued that both the size-explanation and the potentialist explanation rely on the minimal explanation and add nothing to it. I take the problems with these alternative answers to give us reason to reconsider the minimal explanation as an answer to the why-question. But I am also convinced that the core complaint against the minimal explanation stated in Section 3.1 is correct. The why-question asks why the universe of sets is not a set, while the minimal explanation only explains why the universe of sets is not a set according to ZFC; we should thus demand at least some independent reason why the axioms that prohibit the existence of the universal set are there in the first place.

In what follows, I propose a close alternative to the minimal explanation, which I call 'the conception-based explanation'. I argue that the conception-based explanation avoids the core complaint against the minimal explanation at least as well as the size-actualist and potentialist explanations. The conception-based explanation is also more parsimonious than the size-actualist and potentialist explanations, and it doesn't face the problems discussed above in Sections 3.2–3.3. I conclude that the conception-based explanation is the best available answer to the why-question.

In Section 3.3.1, I explained the iterative conception of sets. More generally, and in line with common usage, I understand a "conception" of sets to be a way of thinking about sets that is generally accepted or presupposed within some particular mathematical community. The iterative conception of sets, arguably originating in Zermelo (1930), is a way of thinking about sets as forming the iterative hierarchy of sets described in Section 3.3.1. This way of thinking about sets is at the origin of the why-question; philosophers who ask the why-question assume the iterative conception of sets, as can be seen in the quote from Simmons in Section 3.1 or in the following quote from Linnebo:

> By 'set' I mean set as on the iterative conception, according to which sets are "formed" in stages. (Linnebo, 2010, 144)[36]

Now consider the following response to the why-question:

Conception-Based Explanation The universe of sets is not a set because the supposition that it is contradicts some axioms of ZFC, and these axioms are part of the iterative conception of sets.

Just as the minimal explanation has various instances, so does the conception-based explanation. Here is one such instance, in analogy with Minimal Explanation$_{\mathrm{F,\ P}}$:

Conception-Based Explanation$_{F,\ P}$ The universe of sets is not a set because the supposition that it is contradicts Foundation and Pairing, and Foundation and Pairing are part of the iterative conception of sets.

What the conception-based explanation adds to the minimal explanation is the claim that the axioms that prohibit the existence of a universal set are part of the iterative conception of sets. In particular, what Conception-Based Explanation$_{\mathrm{F,\ P}}$ adds to

[36]See also Studd (2013, 698).

Minimal Explanation$_{\text{F, P}}$ is the claim that Foundation and Pairing are part of the iterative conception of sets. The conception-based explanation is thus a "deep" explanation in the sense of Section 3.1. 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, they are part of what we (who ask the why-question) usually believe or presuppose about sets. If the axioms that prohibit the existence of the universal set are indeed part of the iterative conception of sets, then the conception-based explanation is one we—we who ask the why-question and work with the iterative conception of sets—should be happy with.

What remains to be shown is that at least some axioms that prohibit the existence of a universal set are indeed part of the iterative conception of sets. In other words, one needs to defend either that Foundation and Pairing are part of the iterative conception of sets, or that Separation is part of the iterative conception of sets, etc. But recall that many philosophers—and perhaps most famously Boolos (1971)—have already argued that Foundation and Pairing (and Separation) are part of the iterative conception of sets. Here, for instance, is a summary of the argument for why Pairing is part of the iterative conception of sets. According to the iterative conception of sets, every set appears at some stage. Assume that a and b appear at some stage. Without loss of generality, assume that b appears after a. Then the pair set $\{a, b\}$ appears immediately after the stage where b appears, since, according to the iterative conception, that next stage is the stage where all the sets of sets formed so far appears. Pairing follows.[37]

A similar argument can be given for the claim that Foundation is part of the iterative conception of sets.[38] I will not rehearse it here, but instead note that philosophers have often maintained that Foundation is not only part of but essential to the iterative conception of sets. See for instance Parsons and Boolos:

> One can state in approximately neutral fashion what is essential to the 'iterative' conception: sets form a well-founded hierarchy in which the elements of a set precede the set itself. In axiomatic set theory, this idea is most directly expressed by the axiom of foundation, which says that any non-empty set has an '∈-minimal' element. (Parsons, 1977, 335f.)

> Whatever tenuous hold on the concepts of *set* and *member* were given one by Cantor's definition of "set" and one's ordinary understanding of "element," "set," "collection," etc. is altogether lost if one is to suppose that some sets are members of themselves. (Boolos, 1971, 219)

Thus Foundation and Pairing are part of the iterative conception of sets, as is assumed by Conception-Based Explanation$_{\mathrm{F,\,P}}$. Similar arguments can be given for other instances of the conception-based explanation, but it is enough for my purposes here to argue for the assumptions of just one instance of the conception-based explanation.

Let us now examine how the conception-based explanation fares in comparison with the potentialist or size-actualist explanations specifically with respect to the core complaint against the minimal explanation stated in Section 3.1.

Take the potentialist explanation first. I claim that the potentialist explanation for the why-question is deep in the same way as the conception-based explanation is deep.

[37] This argument can be found for instance in Maddy (1988, 485).
[38] Aside from Boolos (1971), see also for instance Shoenfield (1977, 327).

As we saw in Section 3.3.1, potentialists motivate the axioms for modal set theory from the iterative conception of sets, and then provide a minimal explanation from within modal set theory. So their explanation is fundamentally of the same kind as the conception-based explanation: the principles that prohibit the existence of a universal set in the potentialist explanation are put forward as being motivated or justified by the iterative conception of sets. The problem with the potentialist explanation is that it commends a spurious and costly detour through principles that are either unexplained (such as the principles of modal set theory with the idiosyncratic and primitive notion of modality, seen in Section 3.3.2) or undefended (such as Specifiability, seen in Section 3.3.3). These principles are *much less* obviously part of the iterative conception of sets than Foundation, Pairing, or Separation—unless of course the principles of modal set theory are mere notational variants of ZF, but in which case nothing is added by the potentialist explanation.[39] So the conception-based explanation fares at least as well as the potentialist explanation with respect to the core complaint: If one accepts that the potentialist explanation avoids the core complaint, then one should also accept that the conception-based explanation avoids the core complaint. But the conception-based explanation (i) is more parsimonious than the potentialist explanation, (ii) makes no controversial assumptions concerning what is part of the iterative conception of sets, and (iii) doesn't face the problems explained in Sections 3.3.2–3.3.3.

[39]Potentialists themselves might agree with this point. Indeed, Linnebo himself simply assumes Foundation in his modal set theory (Linnebo, 2013).

Now take the size-explanation. The size-explanation for the why-question appeals to the Limitation of Size Principle, which quite plausibly is not part of the iterative conception of sets. But the Limitation of Size Principle also stems from a particular conception of sets, arguably originating in Cantor's work, according to which sets are "small" collections.[40] Once again, the conception-based explanation thus fares at least as well as the size-explanation with respect to the core complaint: The conception-based explanation and the size-explanation both avoid the core complaint by appealing to what is part of our conception of sets. But the conception-based explanation doesn't face the serious objection raised in Section 3.2.1.

So the conception-based explanation fares at least as well with respect to the core complaint as the size-explanation and the potentialist explanation: the size-explanation, the potentialist explanation, and the conception-based explanation all avoid the core complaint against the minimal explanation by appealing to what is part of our conception of sets. But the conception-based explanation is more parsimonious and less problematic than either the size-explanation and the potentialist explanation. The conception-based explanation should thus be preferred to both the size-explanation and the potentialist explanation; it is the best available answer to the why-question.

We saw that everyone in the debate over the why-question presupposes the iterative conception. We also saw that size-actualists and potentialists alike appeal to what is part

[40] For more on the conception of sets based on the Limitation of Size Principle, see for instance Maddy (1988, 484f.) or Hallett (1984).

of our conception of sets in explaining why the universe of sets is not a set. Hence the conception-based explanation provides a satisfactory response to the why-question by the lights of everyone involved in the debate. This concludes my case for the superiority of the conception-based explanation over its rivals.

At some point, we may want to ask where the iterative conception itself comes from, and how it, in turn, is justified. The answer to these questions could then also provide us with a better understanding of the nature of the conception-based explanation. I will conclude the paper by providing a roadmap for answering these further questions.

In my view, the iterative conception of sets is part of the (or at least *a*) concept of set; it is part of what we—we who ask the why-question and work with the iterative conception—mean by 'set'.[41] 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; they are part of what we—we who ask the why-question and work with the iterative conception—mean by 'set'. This view is often at least implicitly shared among participants in the present dispute. For instance, at least on the face of it, Linnebo in the previous quote states that the iterative conception is part of what we mean by 'set'. Also in his last quote above, Boolos seems to be stating that if we know anything about the concepts expressed by 'set' and 'member'—at least the concepts expressed by 'set' and 'member' on Cantor's usage of these expressions—then we know that sets cannot be members of themselves, which is an elementary consequence of Foundation. See for

[41] For simplicity, I assume here that the concept of set is the meaning of our expression 'set'.

instance also Burgess concerning Separation:

> However formulated, the assumption of separation is so fundamental to Cantorian thought that it is arguably inappropriate to apply Cantor's word 'set' [*Menge*] to theories (such as Quine's NF and ML) that do not accept it. In other words, separation may be regarded as a partial explication of the concept of set, indicating what sets are supposed to be like if they exist. (Burgess, 2004, 203)

It is beyond the scope of this paper to argue that the iterative conception—or more specifically that Foundation, Pairing, and Separation—are conceptual truths, especially given that the existence of conceptual truths is a highly disputed philosophical claim.[42] As I explained above, we also don't need to argue for this in order to show that the conception-based explanation is the best available answer to the why-question. What I do want to point out here, however, is that if the iterative conception of sets is indeed conceptually true, then the conception-based explanation should be understood as a kind of *conceptual* explanation. And this would show that the conception-based explanation is the best possible explanation for why there is no set of all sets.

Let me explain this with a somewhat simplistic analogy. Consider the question: "Why can't one be a happily married bachelor?" The equivalent "minimal" explanation here would be: "One can't be a happily married bachelor, because one would then be both married and a bachelor, and bachelors are unmarried. Contradiction." Now consider the following conceptual explanation that goes slightly beyond this minimal explanation,

[42] There is a long line of arguments against the existence of conceptual or analytic truths and/or of their epistemological consequences, including Quine (1951), Putnam (1962) and the most recent Williamson (2007).

in that it makes explicit that the facts to which this minimal explanation appeals are conceptual truths: "One can't be a happily married bachelor because one would then be both married and a bachelor, and bachelors are by definition unmarried (or, it is a conceptual truth that bachelors are unmarried). Contradiction." The conceptual explanation here is clearly more satisfactory than the minimal explanation; it gives some independent reason for why the assumptions that prohibit bachelors from being happily married are there in the first place. It is also clearly a satisfactory answer to the given question: nothing more is needed to explain why one cannot be a happily married bachelor once one knows the definition of 'bachelor'. Now if the iterative conception—or Foundation, Pairing, and Separation—are conceptually true, then the conception-based explanation should be understood to be of the same kind as this conceptual explanation: the fact that there is no universe of all sets is an elementary consequence of conceptual truths about sets, and that is as good an explanation of why the universe of sets is not a set as we can get.

So if the conception-based explanation is indeed a conceptual explanation, we can see why the conception-based explanation is the best answer to the why-question. Of course, much more needs to be said to defend the view that the iterative conception of sets is conceptually true. The bachelor example above only serves to illustrate the point that a conceptual explanation would be all one could ask for; I don't want to suggest that the mathematical case is quite as simple. In order to defend that the iterative conception of sets is conceptually true, one would need to examine, just for one example, how exactly this view fares with the existence of alternative set theories. I leave the question of how exactly the conceptual explanation would look like for future research.

References

Ayer, A. J. (1936). *Language, Truth and Logic*. London: Victor Gollancz Ltd.

Bagaria, J., Koellner, P., & Woodin, H. (ms). Large cardinals beyond choice. Manuscript.

Bealer, G. (1998). A theory of concepts and concept possession. *Philosophical Issues*, 9, 261–301.

Bealer, G. (1999). A theory of the a priori. In J. Tomberlin (Ed.), *Philosophical Perspectives 13: Epistemology* (pp. 29–55). Oxford: Blackwell.

Benacerraf, P. & Putnam, H., Eds. (1983). *Philosophy of Mathematics*. Cambridge: Cambridge University Press, 2nd edition.

Bengson, J. (2016). Practical perception and intelligent action. *Philosophical Issues*, 26(1), 25–58.

Bengson, J. & Moffett, M. A. (2007). Know-how and concept possession. *Philosophical Studies*, 136, 31–57.

Boghossian, P. (1996). Analyticity reconsidered. *Noûs*, 30(3), 360–391.

Boghossian, P. (2003a). Blind reasoning. *Aristotelian Society, Supplementary Volume*, 77, 225–248.

Boghossian, P. (2003b). Epistemic analyticity: A defense. *Grazer Philosophische Studien*, 66(1), 13–35.

Boolos, G. (1971). The iterative conception of set. *Journal of Philosophy*, 68, 215–231.

Boolos, G. (1984). To be is to be the value of a variable (or to be some values of some variables). *Journal of Philosophy*, 81, 430–449.

Burge, T. (1979). Individualism and the mental. In P. French, Theodore E. Uehling, Jr, & H. K. Wettstein (Eds.), *Midwest Studies in Philosophy IV: Studies in Metaphysics* (pp. 73–121). Minneapolis, MN: University of Minnesota Press.

Burge, T. (1986). Intellectual norms and foundations of mind. *Journal of Philosophy*, 83(12), 697–720.

Burge, T. (1993). Concepts, definitions, and meaning. *Metaphilosophy*, 24(4), 309–325.

Burgess, J. P. (2004). E pluribus unum: Plural logic and set theory. *Philosophia Mathematica*, 12(3), 193–221.

Cantor, G. (1899). Letter to Dedekind. In J. van Heijenoort (Ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Harvard University Press.

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.

Carnap, R. (1937). *The Logical Syntax of Language*. London: Routledge & Kegan Paul.

Carnap, R. (1939). Foundations of logic and mathematics. *International Encyclopedia of Unified Science*, 1(3).

Carnap, R. (1950a). Empiricism, semantics and ontology. *Revue International de Philosophie*, 4, 20–40. Reprinted in 1956, *Meaning and Necessity: A Study in Semantics and Modal Logic*, 2nd edn. University of Chicago Press, Chicago.

Carnap, R. (1950b). *Logical Foundations of Probability*. University of Chicago Press.

Carnap, R. (1952a). Meaning postulates. *Philosophical Studies*, 3(5), 65–73.

Carnap, R. (1952b). Quine on analyticity. In R. Creath (Ed.), *Dear Carnap, Dear Van* (pp. 427–432). University of California Press.

Carnap, R. (1963). W. V. Quine on logical truth. In P. A. Schilpp (Ed.), *The Library of Living Philosophers Volume XI: The Philosophy of Rudolf Carnap* (pp. 915–922). Open Court Publishing Co.

Chalmers, D. (2004). Epistemic two-dimensional semantics. *Philosophical Studies*, 118(1/2), 153–226.

Chudnoff, E. (ms). Grounding and entailment. Manuscript.

Cohen, P. (1971). Comments on the foundations of set theory. In D. Scott (Ed.), *Axiomatic Set Theory: Proceedings of Symposia in Pure Mathematics*, volume 13 (pp. 9–15). American Mathematical Society.

Dancy, J. (2004). *Ethics without Principles*. Oxford University Press.

Dummett, M. (1991). *Frege's Philosophy of Mathematics*. Harvard University Press.

Eagle, A. (2008). Mathematics and conceptual analysis. *Synthese*, 161(1), 67–88.

Feferman, S. (1996). Gödel's program for new axioms: why, where, how and what? In P. Hajek (Ed.), *Gödel '96: Logical Foundations of Mathematics, Computer Science and Physics—Kurt Gödel's Legacy*. Springer-Verlag.

Feferman, S. (1999). Does mathematics need new axioms? *The American Mathematical Monthly*, 106(2), 99–111.

Fine, K. (2006). Relatively unrestricted quantification. In A. Rayo & G. Uzquiano (Eds.), *Absolute Generality* (pp. 20–44). Oxford University Press.

Forster, T. E. (1992). *Set Theory with a Universal Set*. Clarendon Press.

Fraenkel, A. A., Bar-Hillel, Y., & Lévy, A. (1973). *Foundations of Set Theory*. North-Holland.

Frege, G. (1980 [1884]). *The Foundations of Arithmetic*. Oxford: Blackwell, 2nd edition.

Gödel, K. (1944). Russell's mathematical logic. In P. Schlipp (Ed.), *The Philosophy of Bertrand Russell, Library of Living Philosophers*, volume 5 (pp. 123–153). Evanston: Northwestern University.

Gödel, K. (1947). What is Cantor's continuum problem? *The American Mathematical Monthly*, 54(9), 515–525.

Gödel, K. (1983 [1964]). What is Cantor's continuum problem? In P. Benacerraf & H. Putnam (Eds.), *Philosophy of Mathematics: Selected Readings* (pp. 470–485). Cambridge University Press.

Gödel, K. (1995). Some basic theorems on the foundations of mathematics and their implications. In S. Feferman, J. Dawson, W. Goldfarb, C. Parsons, & R. Solovay (Eds.), *Kurt Gödel: Collected Works*, volume 3 (pp. 304–323). Oxford University Press.

Greenberg, M. (2014a). Troubles for content I. In A. Burgess & B. Sherman (Eds.), *Metasemantics: New Essays on the Foundations of Meaning* chapter 6. Oxford University Press.

Greenberg, M. (2014b). Troubles for content II. In A. Burgess & B. Sherman (Eds.), *Metasemantics: New Essays on the Foundations of Meaning* chapter 7. Oxford University Press.

Hahn, H. (1959). Logic, mathematics and knowledge of nature. In A. J. Ayer (Ed.), *Logical Positivism* (pp. 147–161). The Free Press.

Hallett, M. (1984). *Cantorian Set Theory and Limitation of Size*. Oxford University Press.

Hamkins, J. D. (2012). The set-theoretic multiverse. *The Review of Symbolis Logic*, 5(4), 419–449.

Hellman, G. (1989). *Mathematics without Numbers*. Oxford: Oxford University Press.

Holmes, R. (1998). *Elementary Set Theory with a Universal Set*, volume 10. Cahiers du Centre de Logique, Academia, Louvain-la-Neuve.

Holmes, R. (2001). Strong axioms of infinity in NFU. *Journal of Symbolic Logic*, 66(1), 87–116.

Incurvati, L. (2012). How to be a minimalist about sets. *Philosophical Studies*, 159(1), 69–87.

Jackson, F. (1998a). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.

Jackson, F. (1998b). Reference and description revisited. *Noûs*, 32, 201–218.

Jané, I. (1995). The role of the absolute infinite in Cantor's conception of set. *Erkenntnis*, 42(3), 375–402.

Jensen, R. B. (1969). On the consistency of a slight (?) modification of Quine's NF. *Synthese*, 19, 250–263.

Kipper, J. (2012). *A Two-Dimensionalist Guide to Conceptual Analysis*. Ontos Verlag.

Koellner, P. (2006). On the question of absolute undecidability. *Philosophia Mathematica*, 14(2), 153–188.

Koellner, P. (2009). On reflection principles. *Annals of Pure and Applied Logic*, 157(2–3), 206–219.

Koellner, P. (2011). Independence and large cardinals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Summer 2011 edition.

Koellner, P. (msa). Actualism and potentialism. Manuscript.

Koellner, P. (msb). Hamkins on the multiverse. Manuscript.

Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), *Semantics of Natural Language* (pp. 253–355, 763–9). Dordrecht: D. Reidel. Revised edition published in 1980 as *Naming and Necessity* (Harvard University Press, Cambridge, MA).

Kunen, K. (1980). *Set Theory: An Introduction to Independence Proofs*. North-Holland.

Lévy, A. (1976). The role of classes in set theory. In G. H. Müller (Ed.), *Sets and Classes: On the Work by Paul Bernays* (pp. 173–216). North-Holland.

Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61, 343–77.

Lewis, D. (1991). *Parts of Classes*. Oxford: Blackwell.

Lewis, D. (1995). Reduction of mind. In S. Guttenplan (Ed.), *The Blackwell Companion to the Philosophy of Mind* (pp. 412–431). Oxford: Oxford University Press.

Linnebo, Ø. (2010). Pluralities and sets. *Journal of Philosophy*, 107(3), 144–164.

Linnebo, Ø. (2013). The potential hierarchy of sets. *Review of Symbolic Logic*, 6(2), 205–228.

Machery, E. (2009). *Doing without Concepts*. Oxford University Press.

Maddy, P. (1988). Believing the axioms. I. *The Journal of Symbolic Logic*, 53(2), 481–511.

Maddy, P. (2011). *Defending the Axioms: On the Philosophical Foundations of Set Theory*. Oxford University Press.

Martin, D. A. (2005). Gödel's conceptual realism. *The Bulletin of Symbolic Logic*, 11(2), 207–224.

Martin, D. A. (ms). Completeness or incompleteness of basic mathematical concepts.

Mayberry, J. (1986). Review of Hallett (1984). *Philosophical Quarterly*, 36(144), 429–434.

Menzel, C. (2014). Wide sets, ZFCU, and the iterative conception. *Journal of Philosophy*, 111(2), 57–83.

Parsons, C. (1974). Sets and classes. *Noûs*, 8(1), 1–12.

Parsons, C. (1977). What is the iterative conception of set? In R. E. Butts & J. Hintikka (Eds.), *Proceedings of the 5th International Congress of Logic, Methodology and Philosophy of Science 1975, Part I: Logic, Foundations of Mathematics, and Computability Theory* (pp. 335–367). Reidel. Reprinted in (Benacerraf & Putnam, 1983).

Parsons, C. (1983). Sets and modality. In *Mathematics in Philosophy* (pp. 298–341). Cornell University Press.

Parsons, C. (2014). Analyticity for realists. In J. Kennedy (Ed.), *Interpreting Gödel: Critical Essays*. Cambridge University Press.

Pavese, C. (2015). Practical senses. *Philosopher's Imprint*, 15(29), 1–25.

Peacocke, C. (1992). *A Study of Concepts*. MIT Press.

Peacocke, C. (1998). Implicit conceptions, understanding and rationality. *Philosophical Issues*, 9, 48–88.

Peacocke, C. (2000). Explaining the a priori: The programme of moderate rationalism. In P. Boghossian & C. Peacocke (Eds.), *New Essays on the A Priori* (pp. 225–285). Oxford University Press.

Peacocke, C. (2008). *Truly Understood*. Oxford University Press.

Priest, G. (1979). The logic of paradox. *Journal of Philosophical Logic*, 8, 219–241.

Priest, G. (1995). *Beyond the Limits of Thought*. Cambridge University Press.

Prinz, J. J. (2002). *Furnishing the Mind*. MIT Press.

Putnam, H. (1962). The analytic and the synthetic. In H. Feigl & G. Maxwell (Eds.), *Minnesota Studies in the Philosophy of Science*, volume III. University of Minnesota Press.

Putnam, H. (1970). Is semantics possible? *Metaphilosophy*, 1(3), 187–201.

Putnam, H. (1975a). The meaning of 'meaning'. In K. Gunderson (Ed.), *Language, Mind and Knowledge*, volume 7 of *Minnesota Studies in the Philosophy of Science* (pp. 131–193). Minneapolis: University of Minnesota Press.

Putnam, H. (1975b). *Mind, Language and Reality: Philosophical Papers, volume 2*. Cambridge: Cambridge University Press.

Quine, W. V. (1951). Two dogmas of empiricism. *Philosophical Review*, 60, 20–43.

Rayo, A. & Linnebo, Ø. (2012). Hierarchies ontological and ideological. *Mind*, 121(482), 269–308.

Recanati, F. (2000). Deferential concepts: A response to woodfield. *Mind and Language*, 15(4), 452–464.

Rey, G. (1998). What implicit conceptions are unlikely to do. *Philosophical Issues*, 9, 93–104.

Rosen, G. (2010). Metaphysical dependence: Grounding and reduction. In B. Hale & A. Hoffmann (Eds.), *Modality: Metaphysics, Logic, and Epistemology* (pp. 109–136). Oxford: Oxford University Press.

Russell, G. (2008). *Truth in Virtue of Meaning*. Oxford University Press.

Schwarz, W. (2014). Against magnetism. *Australasian Journal of Philosophy*, 92(1), 17–36.

Shoenfield, J. R. (1977). Axioms of set theory. In J. Barwise (Ed.), *Handbook of Mathematical Logic* (pp. 321–344). North-Holland.

Simmons, K. (2000). Sets, classes and extensions: A singularity approach to Russell's paradox. *Philosophical Studies*, 100(2), 109–149.

Smith, S. R. (2015). Incomplete understanding of concepts: The case of the derivative. *Mind*, 124(496), 1163–1199.

Studd, J. (2013). The iterative conception of set. *Journal of Philosophical Logic*, 42(5), 695–725.

Tait, W. (2005). Constructing cardinals from below. In *The Provenance of Pure Reason* chapter 6, (pp. 133–154). Oxford University Press.

Tall, D., Ed. (1991). *Advanced Mathematical Thinking*. Kluwer Academic Publishers.

Uzquiano, G. (2003). Plural quantification and classes. *Philosophia Mathematica*, 11(1), 67–81.

Welch, P. (ms). Conceptualism: Sets and classes. Manuscript.

Williamson, T. (2006). Conceptual truth. *Aristotelian Society, Supplementary Volume*, 80(1), 1–41.

Williamson, T. (2007). *The Philosophy of Philosophy*. Malden, MA: Blackwell.

Williamson, T. (2016). Absolute provability and safe knowledge of axioms. In L. Horstein & P. Welch (Eds.), *The Limits of Mathematical Knowledge* (pp. 243–253). Oxford University Press.

Yablo, S. (2004). Circularity and paradox. In V. F. H. Thomas Bolander & S. A. Pedersen (Eds.), *Self-Reference*. CSLI.

Zermelo, E. (1930). Über Grenzzahlen und Mengenbereiche. *Fundamenta Mathematica*, 16, 29–47.

Zilber, B. (ms). Axioms of set theory. Manuscript.