A Benacerraf Problem for Higher-Order Metaphysics

Abstract

Higher-order metaphysics is in full swing. Its proponents argue that higher-order logic should replace set theory at the foundations of mathematics and metaphysics. But amid the enthusiasm, surprisingly little attention has been paid to some serious epistemological challenges facing the program - foremost among them a variant of the Benacerraf challenge, developed by Field and Clarke-Doane. Roughly put, the challenge is to explain the reliability of our higher-order logical beliefs. A similar problem is familiar from the philosophy of set theory, where it has led to a pluralist reconception of the foundations of mathematics. In this paper, I argue that regardless of whether higher-order logic is preferable to set theory on abductive grounds, they stand or fall together when faced with this epistemological challenge. They are companions in guilt (or innocence). I conclude that, absent other solutions, a promising path forward is to adopt a pluralist approach to higher-order logic. The consequences of such a shift are difficult to overstate.

1 Higher-Order Metaphysics

Higher-order logic, of the kind currently popular in metaphysics, extends first-order logic by allowing quantification not only over individual variables but also over predicates, relations, and functions of various types. This means that in addition to statements about objects, higher-order logic can express complex claims about properties of properties, relations between relations, and functions of functions, resulting in a more expressive framework for formal reasoning. It is particularly powerful in capturing the semantics of mathematical and metaphysical theories, as it can encode notions such as class membership, cardinality, and metaphysical necessity.[1]

Its proponents contend that higher-order logic should replace set theory at the foundations of mathematics and metaphysics. Without going into too much detail, I take this to mean that it should be used to:

- Interpret our canonical mathematical and metaphysical theories (broadly

[1]See Williamson [2013], Bacon and Dorr [2024], and Goodsell and Yli-Vakkuri [MS] for more formal details.

construed), such as our theories of arithmetic, classes, syntax, semantics, modality, and so on;[2]

- Elucidate our informal or imprecise concepts, such as those of numbers, properties, propositions, absolute necessity, and so on;[3]
- Serve as the benchmark for acceptable theories;[4]
- Provide a precise account of a rigorous proof or argument;[5]
- Provide a common arena in which to carry out all of our mathematical and metaphysical theorizing.

I take no stand in this paper on whether we should adopt higher-order logic as our foundational meta-theory instead of set theory. Perhaps it fares better than set theory in an overall abductive comparison. Perhaps not. My focus will be on the claim that even if it does, higher-order logic fares no better than set theory with respect to a variant of the Benacerraf challenge due to Field and Clarke-Doane. Set theory and higher-order logic are companions in guilt (or innocence) with respect to this challenge. So the challenge cannot be avoided simply by adopting higher-order logic in place of set theory.

2 The Benacerraf Challenge

Benacerraf wrote, in his influential *"Mathematical Truth"*:

> [O]n a realist (i.e., standard) account of mathematical truth our explanation of how we know the basic postulates must be suitably connected with how we interpret the referential apparatus of the theory. [...] [But] what is missing is precisely [...] an account of the link between our cognitive faculties and the objects known.[6]

He worried that there is no such account. As such, he thought, we must either give up the "standard" realist interpretation of mathematics or settle for an epistemic mystery. This worry stemmed from his commitment to the causal theory of knowledge. He wrote:

> I favor a causal account of knowledge on which for X to know that S is true requires some causal relation to obtain between X and the referents of the names, predicates, and quantifiers of S.[.. . .] [But] [...] combining this view of knowledge with the "standard" view

[2] As, for example, Goodsell [2021] does for arithmetic.
[3] As, for example, Bacon [2018] does with absolute necessity.
[4] Every acceptable theory has to be consistent according to the logic.
[5] As Goodsell and Yli-Vakkuri [MS] put it, 'deductive validity according to LF is to be used as the criterion for assessing what follows from the verdicts, hypotheses, or conjectures of any science'.
[6] Benacerraf [1973].

> of mathematical truth makes it difficult to see how mathematical knowledge is possible.[7]

Now, we might immediately think that even if this is a problem for set theory, with its vast ontology of abstract objects, it is not a problem for higher-order logic, which does not come along with any distinctive first-order ontology, abstract or otherwise. However, the causal theory of knowledge has since been rejected for a variety of independent reasons. As such, Benacerraf's specific problem has largely been put aside. Nonetheless, it is widely agreed that he was on to a significant challenge for mathematical realism. The genuine issue to which Benacerraf was pointing is commonly thought to have been identified by Field, and precisified by Clarke-Doane.

There are four main steps in Field's account:[8]

1. We grant the truth of our mathematical beliefs.
2. We grant that we are defeasibly justified in holding these beliefs. This is significant: at this stage, we are not trying to justify the beliefs. Rather, we are assuming they are already justified.
3. We must then explain the connection between our beliefs and the truths, freely relying on the assumptions that our beliefs are true and defeasibly justified. We must show that the method by which we formed these beliefs is reliable - that the truth of our beliefs is not merely a matter of luck.
4. If we have reason to think this cannot be done, our beliefs are undermined. If we cannot explain how the process that led us to these beliefs reliably tracks the truth, then we ought to give them up.

According to Field, the challenge is to show that there is some reason to think we can explain the reliability of our mathematical beliefs, given that we are allowed to assume these beliefs are both true and defeasibly justified. Field holds that if we cannot meet this explanatory demand, our beliefs are undermined. Clarke-Doane further elaborates on the nature of the challenge.[9] He observes that there must be a sense of 'explain the reliability' such that, first, it seems impossible for us to explain the reliability of our mathematical beliefs, and second, this failure undermines those beliefs. He suggests that *safety* can fulfill this role.

> A belief that p, is safe just in case, we could not easily have had a false belief as to whether p, using the method that we actually used to determine whether p.

[7]Benacerraf [1973].
[8]Field [1989].
[9]Clarke-Doane [2020 (a)].

According to Clarke-Doane, there is a good case to be made both that we cannot account for the safety (of a least a significant portion) of our mathematical beliefs, and that recognizing this undermines those beliefs.[10]

So, the version of the Benacerraf challenge for a domain D with which I am concerned in this paper is the following: to show, for each of our D-beliefs, while freely assuming that they are defeasibly justified and true, that we could not easily have come to hold false such beliefs using the method we actually used to form them.

Within higher-order metaphysics the general epistemological method appears to be an abductive one, according to which we try to choose the most theoretically virtuous higher-order logic which allows us to prove everything that we need to prove in all of science. Suppose that higher-order metaphysicians settle on a specific higher-order logic L using this methodology, just as set theorists have settled on ZFC. The challenge is then to explain how we could not easily have settled on false logical beliefs using this method, freely using the assumptions that L is true and that our belief in it is defeasibly justified on abductive grounds.

This version has a number of virtues. First, it is agnostic with respect to different theories of knowledge. In particular, it does not depend on the casual theory of knowledge.

Second, it is not merely a skeptical challenge. It is not the challenge to show that our logical beliefs are safe. That would require showing that they are true. Rather, it is the challenge to show that our beliefs are safe while freely making use of the assumption that they are true and defeasibly justified. This can clearly be met, even if beliefs turn out not to be safe. Consider, by way of illustration, that we seem able to meet the challenge for our perceptual beliefs. Assuming they are true and defeasibly justified, we can point to a complex neurobiological account of how we track the physical truths about our environment, alongside an evolutionary account of how we came to possess faculties that enable such tracking. Yet none of this builds in the actual safety of our perceptual beliefs. All of it is consistent with the possibility that those beliefs are false - for example, if we are brains in vats.

Third, this formulation does not depend on assuming that higher-order logic involves a peculiar abstract ontology. It depends only on the view that logical truths are mind-and-language-independent. This is reflected in the wide applicability of the challenge. It closely resembles a debunking argument. Such arguments should be familiar from the case of ethics, where they have motivated many metaethicists to adopt anti-realism, as well as consciousness, aesthetics, validity, and modality, where there is no obvious commitment to a distinctive first-order ontology.[11][12]

Fourth, whereas Benacerraf's version of the problem targeted all of our mathematical beliefs, this version allows for a more fine-grained application. It may

[10] See Clarke-Doane [2020 (a)] for more details.

[11] See for instance, Street [2006], Chalmers [2020], and Schechter [2010].

[12] These first three features of this version are originally discussed in by Clarke-Doane [2020 (a)].

be that we can account for the safety (again, assuming defeasible justification and truth) of some logical beliefs but not others. Consider the case of set theory. We might think we can do this for the axioms of ZFC, and that classical first-order consequence preserves this feature. In that case, we could meet the challenge for our belief in ZFC, but not for any set-theoretic beliefs that go beyond it. Indeed, it seems plausible that this kind of reasoning has motivated Joel Hamkins's set-theoretic multiverse view, according to which there are many distinct set-theoretic universes. These universes agree on the theorems of ZFC but disagree on statements like the Continuum Hypothesis, which are independent of ZFC.[13]

And finally, there are two prominent interpretations of what it means to "explain the reliability" of our beliefs, aside from the safety interpretation: the sensitivity interpretation and the explanation interpretation. A belief that p, is sensitive just in case, had p been different, our belief would have been correspondingly different. According to the sensitivity interpretation, we must explain how our beliefs are sensitive, assuming they are both true and defeasibly justified. The problem is that logical truths are necessary. Given this, it is trivially the case that our logical beliefs are sensitive. The safety interpretation neatly sidesteps this issue.

According to the explanation interpretation, we must show — again assuming that p is true and defeasibly justified — that p's truth figures in the best explanation of our belief that p. But this faces difficulties when applied to logical beliefs. Assuming our logic is true, logical truths will figure in any explanation we offer. Since all logical truths follow from any such explanation, they are trivially part of every explanation. This makes this interpretation particularly unsuitable for application to higher-order logic. Once again, the safety interpretation avoids this problem.

Many philosophers of mathematics have found this challenge deeply troubling in the case of set theory. Indeed, pessimism about our ability to meet it, assuming the standard picture of set theory – that it is about the unique, mind-independent, Platonically existing universe of sets – has motivated some to adopt radical alternatives, including various forms of pluralism.[14] This concern is not unique to mathematics. Similar debunking arguments have pushed philosophers toward anti-realism in domains such as consciousness, morality, and modality.

With this said, I would like to emphasize a few key points. While I agree with those who believe the challenge cannot be met for a significant portion of set theory (given the standard view), I will neither argue for nor assume that conclusion. Likewise, while I agree with philosophers such as Field and Clarke-Doane that failing to meet the challenge for certain beliefs may undermine those beliefs, I will not argue for or assume that claim either. In particular, I do not argue that if the Benacerraf challenge cannot be met for part of higher-order logic, then beliefs in those logical truths are undermined and should be aban-

[13]Hamkins [2012].

[14]See, for example, Balaguer [1995], Linsky and Zalta [1995], and Hamkins [2012].

doned. These broader implications have been addressed at length by others.[15] My concern is only with the conditional: if we cannot meet the challenge for set theory, then we cannot meet it for higher-order logic.

And lastly, the companions-in-guilt thesis is not intended as an argument against higher-order logic in favor of set theory, nor as a claim that the two stand on equal footing when it comes to overall theory choice. Indeed, the thesis is compatible with the idea that higher-order logic fares much better than set theory as a foundation for metaphysics (a point I elaborate in Section 4.6). The point is rather that, even if higher-order logic is superior in this regard, it inherits the epistemological problem — if it is a problem — from set theory.

3 Companions in Guilt

In this section, I make a prima facie case that if the Benacerraf challenge cannot be met for some set-theoretic claims under the standard view — according to which there exists a unique, Platonically existing universe of sets — then it likewise cannot be met for some logical claims under the standard view, where logical truths are taken to be mind-and-language-independent and objective[16]. The case is quite simple. It proceeds as follows:

1. There are logical analogues of many set-theoretic claims

2. If the Benacerraf challenge cannot be met for some set-theoretic claim p, then we could just as easily arrive, via an overall abductive method, at a set theory that includes p, as a set theory that includes $\neg p$. In other words, mathematics, science, and metaphysics can be done equally well by using set theories that disagree about p.

[15] See, for example, Benacerraf [1973], Field [1989], and Street [2006].

[16] Objectivity is the opposite of pluralism. The idea is that claims about a domain D are objective just in case they have a uniquely good intended interpretation. So set theory is objective if there is just one universe of sets. If there were two (or more) universes of sets, one in which CH holds and one in which it doesn't, then CH would not be objective. It is important here that we think that there are two universes of sets, and as such two equally good interpretations. Not just that there are any two different interpretations, such as different models of set theory. We would still say that set theory is objective if there were just one universe of sets, and a variety of different models of set theory, which give different interpretations. As the universe would be a uniquely good interpretation. Similarly, in the case of ethics, if there is just one uniquely good interpretation of 'good', then it is objective. But if for instance, there are two equally good interpretations, say 'maximizes utility' and 'is what the virtuous person would do', then claims which come apart on these interpretations are not objective. Higher-order logic is objective then if it does not have a variety of equally good interpretations. For instance, suppose all the instances of the schema $\forall^{\sigma\to t}X\forall^{\sigma}x(Xx \vee \neg Xx)$ are true. These are objective if there is no equally legitimate sense in which the instances of $\neg(\forall^{\sigma\to t}X\forall^{\sigma}x(Xx \vee \neg Xx))$ are true. Informally, they are objective if there is just one unrestricted universal quantifier of each type, and that on these intended interpretations, each instance of $\forall X^{\sigma\to t}\forall x^{t}(Xx \vee \neg Xx)$ is true. Rather than there being at least two unrestricted universal quantifiers of each type, which as such are equally good interpretations, such that on one interpretation the instances of $\forall^{\sigma\to t}X\forall^{\sigma}x(Xx \vee \neg Xx)$ are true, and on the other the instances of $\neg(\forall^{\sigma\to t}X\forall^{\sigma}x(Xx \vee \neg Xx))$ are true.

3. If this is so, then mathematics, science, and metaphysics can be done equally well by higher-order logics that disagree about a logical analogue of p. As such, we could just as easily arrive, via the abductive method, at a higher-order logic that includes this logical analogue of p as one that includes its negation.

4. Thus, if the Benacerraf challenge cannot be met for some set-theoretic claim p, it likewise cannot be met for its higher-order logical analogue.

1: There are higher-order-logical analogues of many set-theoretic claims, including the axiom of pairs, the axiom of choice, and the continuum hypothesis (CH). For example, consider HCH: a formula in the language of pure higher-order logic that ZFC proves is valid (in the model-theoretic sense, assuming the standard semantics) just in case the set-theoretic continuum hypothesis is true. In higher-order logic, we can define what it means for a property to have cardinality $\aleph_0$, $\aleph_1$, or the cardinality of the continuum. As such, in higher-order logic we can express that there is no size between $\aleph_0$ and the continuum, by saying that a property has size $\aleph_1$ just in case it has the size of the continuum. So, $HCH := \forall^{e \to t} X(\aleph_1(X) \leftrightarrow \mathit{Continuum}(X))$.[17] We can think of it as a higher-order version of the set-theoretic continuum hypothesis. Similar higher-order formulations exist for many other set-theoretic claims.

2: If there is no in principle explanation of how our beliefs about set-theoretic claims — such as the axiom of pairs, the axiom of choice, or the continuum hypothesis (CH) — are safe (assuming they are true and justified), then mathematics, science, and metaphysics can be done equally well using set theories that disagree about them.[18]

Let's use CH and its higher-order analogue, HCH, as examples. Suppose that we cannot — even in principle — meet the Benacerraf challenge for the continuum hypothesis, on the assumption that CH is about a unique universe of sets. That is, we cannot explain how it is that we could not easily have ended up with false beliefs about CH, using the very method we actually used — even granting that our beliefs are both true and defeasibly justified.

In particular, suppose our method is an overall abductive one: we adopt the most theoretically virtuous set theory, i.e., the one that best supports mathematics, science, and metaphysics as a whole.[19]

If this method cannot, even in principle, secure the safety of our belief in CH, then it follows that for any set theory S that includes CH as a theorem

[17] See Shapiro [1991] for more details.

[18] As noted earlier, I need not take a stand on whether we can account for the safety of our beliefs about any or all of these set-theoretic claims (on the assumption that they are true and justified). The point is just that if we cannot, then mathematics and science can proceed equally well with theories that disagree about such claims.

[19] I focus on abductive justification throughout this paper, as it seems to be the dominant approach among higher-order metaphysicians. However, I believe that the arguments would apply equally to other epistemological frameworks — e.g., one that appeals to a special faculty of mathematical or logical intuition.

and could be reached via this method, there is an alternative set theory S* that includes ¬CH as a theorem, and which could just as easily have been arrived at using the same abductive approach. And vice versa. Otherwise, we would possess an in principle explanation — again, assuming our belief in CH (or ¬CH) is true and defeasibly justified — of how it is that we could not easily have ended up with false beliefs about CH. Namely: that the abductive method would have led us only to set theories that agree on CH.

3: Consider HCH and $H\neg CH$.[20] The former is valid just in case CH is true; the latter is valid just in case CH is false.[21] Could we, given our supposition about CH, account for how we could have a safe belief about HCH — on the understanding that it is mind-and-language-independent and objective — while assuming that our logic is both true and defeasibly justified?

Our best bet would be to adopt a logic that proves either HCH or $H\neg CH$. That way, we can assume one of them is true and that we are defeasibly justified in believing it, because the logic in question allows us to prove everything we need to prove across all of science. Suppose, then, that we settle on $LF + HCH$.[22] Could we now account for how we could not easily have come to believe a falsehood about HCH? That is, could we rule out that we might have come to believe a logic proving $H\neg CH$, using the same abductive method? There is reason to think not.

I am assuming that there is not even an in principle account of how we could have safe beliefs about CH. And so I am also assuming that mathematics, physics and metaphysics can be done equally well on the assumption that the cardinality of the continuum is the next largest infinite cardinal after a countable infinity, and on the contrary assumption that there are cardinalities between the countable and the continuum. This suggests that if $LF + HCH$ provides all the logical resources we need, then so too does $LF + H\neg CH$. And even if $LF + H\neg CH$ doesn't, it suggests that some other abductively attractive logic which proves $H\neg CH$ will.

If both $LF + HCH$ and a logic proving $H\neg CH$ provide all the logical resources we need for mathematics, physics, and metaphysics, then — even assuming that $LF + HCH$ is true and abductively justified — it remains very difficult to explain how we could not easily have come to believe $H\neg CH$ instead, using the very same abductive method. All we are relying on is choosing a logic that proves what we need to prove — and there are equally good options available which disagree about HCH.[23]

[20] $H\neg CH := \forall^{e\to t}X(\aleph_1(X) \to \neg Continuum(X))$.

[21] Again, see Shapiro [1991] for more details.

[22] See Goodsell and Yli-Vakkuri [MS] for details about the higher-order logic LF.

[23] Recall that I am assuming a meta-set-theoretic and meta-logical stance in which both set theory and higher-order logic are taken to be mind-and-language-independent and objective. This is crucial. To see this, suppose we thought set theory is mind-and-language independent, but regarded higher-order logic as dependent on our linguistic or conceptual practices. We might then fail to meet the challenge for CH — since mathematics, physics, and metaphysics can be done equally well with set theories that disagree about CH — but still succeed in meeting it for HCH, insofar as its truth would depend on our practices, to which we have

4: In short, if there is no in principle account of how we could have safe beliefs about CH — assuming it is both true and defeasibly justified — then there is likewise no in principle account of how we could have safe beliefs about HCH. More generally, and recalling that the continuum hypothesis serves here merely as an illustration, the same reasoning applies to other set-theoretic claims about infinite cardinalities and features of infinite classes.

If we cannot, even in principle, explain how our beliefs about such claims could be safe (again, assuming their truth and defeasible justification), then we likewise cannot explain how our beliefs about their higher-order analogues could be safe. In other words, if the Benacerraf challenge cannot be met for some portion of set theory — on the standard Platonist view — then it also cannot be met for the corresponding portion of higher-order logic, assuming it too is taken to be objective and mind-independent.

4 Objections and Responses

If my prima facie case is correct, set theory and higher-order logic are companions in guilt with respect to the Benacerraf challenge. In this section I will consider and respond to some potential objections.

4.1 Incommensurability: Objection

The case I presented relied on the existence of higher-order analogues of set-theoretic claims. In particular, it hinged on treating HCH as a higher-order analogue of the Continuum Hypothesis. But the only evidence offered for this analogy was a proof in first-order ZFC. That proof, moreover, depends on various trappings of model theory: it defines validity as truth in all models, interprets the higher-order language in set-theoretic terms, and adopts standard (as opposed to Henkin) semantics for higher-order logic. None of this is likely to persuade an audience of higher-order metaphysicians who are, at best, skeptical of ZFC, model theory, and set-theoretic interpretations of higher-order language.

More generally, there may be good reason to think that set theory and higher-order logic are incommensurable in something like the Kuhnian sense.[24] If so, it might not make sense to regard HCH as a higher-order analogue of CH at all, since set-theoretic and higher-order logical claims may not be apt for direct comparison. To the extent that my argument depends on such a comparison being intelligible, it fails.[25]

reliable epistemic access. Similarly, if we thought set theory is objective, but were pluralists about higher-order logic, then again we might fail to meet the challenge for set theory, but meet it for higher-order logic — since both HCH and $H\neg CH$ could be true on different, equally good interpretations of higher-order vocabulary. (See section 5 for more on this point.)

[24] See Kuhn [1962].

[25] Thanks to an anonymous referee for raising this objection.

4.2 Incommensurability: Response

I agree that it is quite reasonable for a higher-order metaphysician to be dubious of the set-theoretic 'proof' that CH is true if and only if HCH is valid. However, a similar bridge claim can be established in a higher-order meta-theory, in which validity is defined in explicitly higher-order terms. Florio and Incurvati show that in a meta-theory which contains the axioms of full-second order logic, as well as those of second-order ZFC, it can be proved that CH is true just in case HCH is a higher-order validity.[26] Thus, even from a higher-order perspective that avoids model theory and set-theoretic interpretations of higher-order language, there is good reason to view HCH as a higher-order analogue of CH. And this point generalizes to many other pairs of set-theoretic and higher-order claims.

Moreover, even if we grant that set theory and higher-order logic are incommensurable, this need not undermine the prima facie case. If we accept that general meta-theories can be incommensurable, then we must abandon the idea of absolute comparisons between theories. What remains is how things appear from within each meta-theoretical perspective. From the standpoint of first-order ZFC set theory, we can prove that CH is true if and only if HCH is valid (in the model-theoretic sense). From the standpoint of higher-order logic, we can likewise prove that CH is true if and only if HCH is valid (in the higher-order sense). Thus, from each relevant perspective, there is good reason to treat HCH as a higher-order analogue of CH. Again, the same holds for many other pairings of set-theoretic and higher-order claims.

4.3 Ontology: Objection

Benacerraf's original presentation of the problem for set theory centered on the postulation of abstract mathematical entities. His concern lay with the metaphysically cordoned-off realm of pure sets: how could we ever come to know about such objects, given that we never causally — or otherwise —interact with them? Moreover, set theory postulates a staggering multitude of such entities, and makes correspondingly precise claims about their structure.

By contrast, higher-order logic does not appear to introduce a distinctive first-order ontology. It makes no ontological commitments beyond the existence of infinitely many objects.[27] The source of the worry in the set-theoretic case is the apparent impossibility of explaining how we could interact with a realm of abstracta. Since higher-order logic seemingly avoids such commitments, it appears not to face the same problem.

[26] The idea is outlined in Florio and Incurvati [2019]. If we let our object language be $l^2_\in$ – a second-order language whose only non-logical predicate is $\in$, and let ϕ^U be the restriction of ϕ's quantifiers to U, then ϕ is a higher-order validity just in case for every non-empty property U and relation E, $\phi[E/\in]^U$ holds (where $\phi[E/\in]$ stands for the metalinguistic formula resulting from replacing all occurrences of $\in$ in ϕ with E.) In other words, higher-order validity is truth with respect to any higher-order domain and any higher-order interpretation of the membership predicate.

[27] Indeed, several prominent higher-order logics, such as Classicism, do not even require this.

4.4 Ontology: Response

This objection misses the point of the variant of the Benacerraf problem that concerns me in this paper. It conflates the access problem with the reliability problem. Even if higher-order logic avoids a burdensome ontology, this does not resolve the issue of epistemic reliability. Supposing that the truth of higher-order logical claims is mind-independent and objective, it remains mysterious how we could reliably determine whether HCH is true, for instance — just as it is mysterious how we could reliably determine the truth of CH in pure set theory.

To make the point vivid, consider analogies with other domains: morality, aesthetics, and consciousness. Even if one is a Quinean nominalist about these, the reliability challenge persists. With realism in the background, the central target of the Benacerraf challenge is, as Kreisel put it, *objectivity, not objects*.

Moreover, it is far from clear that higher-order logic is ontologically innocent. Admittedly, it does not posit a special class of abstract first-order entities. But if we take the higher-order project seriously, then the Quinean dictum —*to be is to be the value of a bound variable*— may naturally extend to higher orders.[28] If higher-order quantification is just as metaphysically perspicuous as first-order quantification, then we may be committed to a range of higher-order entities. For example, to $be_{e \to t}$ would be to be the value of a bound second-order variable. In this case, higher-order logic does bring with it a host of abstruse higher-order entities.[29]

4.5 Justification: Objection

Even granting the foregoing points, there remains a relevant difference between set theory and higher-order logic. Higher-order logic is continuous with the rest of science, and as such, it is apt for empirical justification. As Church, Tarski, Field, and others have shown, there is strong reason to believe that our best scientific theories can be formulated in the language of higher-order logic.[30] These theories face the tribunal of experience holistically, in the Quinean sense. Thus, the empirical success of our best physical theories confers support not only on their specific content, but also on the logic in which they are formulated. On this view, higher-order logic earns its justification in much the same way as our theories of particle physics or space-time geometry do.

In contrast, there is good reason to think that set theory is not continuous with the rest of science, and therefore not apt for empirical justification in the same way. But even if that is mistaken — even if set theory is somehow empirically justifiable — higher-order logic fares better on abductive grounds.[31]

[28] See Quine [1948].

[29] This is similar to the view developed by Trueman [2020].

[30] See, for instance, Church [1940], Tarski [1943] and [1956], Field [1980], Arntzenius and Dorr [2012], and Williamson [2013].

[31] There are many reasons to think this. For instance, we might believe that we need an unrestricted first-order class comprehension axiom, or that we need to be able to do semantics in the way that Tarski did.

Either way, higher-order logic is in a stronger justificatory position. And so it is not susceptible to the Benacerraf challenge in the same way set theory is — since it is precisely the lack of justification for set-theoretic claims that renders them unsafe.

4.6 Justification: Response

There are several points worth making here. First, it is unclear why higher-order logic is supposed to be continuous with the rest of science in a way that set theory is not — especially given, as I noted in Section 4.4, that higher-order logic appears to carry with it a host of abstruse higher-order entities. Second, it is also unclear why set theory could not receive empirical justification via the role it plays in our best scientific theories. On this point, I side with Quine and Lewis. And third, it is not obvious that set theory fares worse than higher-order logic in an overall abductive comparison. Despite the ambitions of some higher-order metaphysicians, the majority of mathematicians, logicians, physicists, and even metaphysicians still seem to prefer ZFC set theory as the framework in which to formulate their theories.

But even if I concede all of these contentious points, granting for the sake of argument that higher-order logic is abductively justified in a way that set theory is not, this does not undermine my argument. Meeting the Benacerraf challenge for a domain D – giving an account of the safety of our D-belief assuming their truth and justification – is independent of whether those beliefs are in fact justified. Suppose, for example, that we have an answer to the challenge for set theory. This could be so even if our set-theoretic beliefs are not in fact justified, since the challenge allows us to assume that they are. Conversely, our beliefs could be justified—for instance, because they enable us to prove everything needed in science—yet we might still fail to meet the challenge, because other incompatible set theories would do just as well in that respect.

So even if we grant the set theorists that their beliefs are justified, the Benacerraf worry still arises. This is crystal clear, given that the challenge allows them to assume that their beliefs are true and justified when attempting to explain safety. Thus, the worry arises for a domain even when our beliefs are in fact justified, and we get to assume as much.

And so, even if our higher-order beliefs are justified on abductive grounds, and we are allowed to assume that our beliefs are justified on these grounds, this does not settle the Benacerraf worry. We still have to explain how it is, making use of the assumptions that our beliefs are true and defeasibly justified, that we could not easily have had false logical beliefs. And my argument has been that if mathematics, physics, and metaphysics can be carried out equally well using a variety of set theories that disagree on key claims, such as the size of the continuum, then there is reason to think that the same is true for higher-order logic. That is, there will be higher-order logics that disagree about higher-order versions of these claims, and yet allow us to prove everything needed in science. And so, the method of choosing the logic that allows us to do everything we want to in science will not reliably lead us to one that gets such claims right.

4.7 Determinacy: Objection

The language of set theory is notoriously indeterminate. It is well known that ZFC admits a wide range of non-standard models, constructed via standard model-theoretic techniques. Famously, Putnam drew on such constructions to argue that the language of set theory is hopelessly indeterminate.[32] In stark contrast, standard models of higher-order logic are categorical. That is, under standard semantics, higher-order logic has only one model up to isomorphism (modulo the number of individuals it posits). Its language admits no non-standard interpretations. These are familiar meta-logical results.

This contrast matters, as it is precisely the indeterminacy of set-theoretic language that undermines the safety of our judgments in that domain. On this view, CH is vulnerable to the Benacerraf challenge because it is formulated in an indeterminate language, whereas HCH, formulated in the determinate language of higher-order logic, is not. As such, the former is epistemically problematic in a way the latter is not.

4.8 Determinacy: Response

This metasemantic point is irrelevant to the present concern, even if true, which I shall question momentarily. While it is often observed that the metasemantic and epistemological worries are motivated by similar considerations, they are importantly distinct. For example, Benacerraf's own [1973] formulation of the challenge appeals to a causal account both of mathematical knowledge and of the determinacy of mathematical language. The same lack of causal interaction underwrites both his metasemantic and epistemological concerns. But despite their shared motivation, these are strictly different worries.

The reliability challenge is an epistemological challenge, not a metasemantic one. The metasemantic worry is that there may be nothing in our linguistic or social practices, or in the world, that makes it the case that we mean *that* $2^{\aleph_0} = \aleph_1$ by 'CH'. This is the concern raised by Putnam in his influential [1980]. That kind of problem may be addressed by a theory of reference, such as that proposed by Lewis [1984]. But the epistemological challenge at issue in this paper is different. Even if we suppose that CH is true, and even that we are defeasibly justified in believing it, we seem unable to show that we could not easily have believed $\neg$CH instead. This doesn't build in any specific assumptions about determinacy, or other semantic notions more generally, other than truth.[33]

Indeed, even if we grant that CH is determinately true, the epistemological worry remains. All that is required for the problem to arise is that it be coherent to suppose that we could easily have come to accept a set theory which proves $\neg$CH (where set theories are individuated syntactically), even while assuming

[32] See Putnam [1980].

[33] This is important from a dialectical perspective because when it comes to truth specifically, you can have a version of Tarski's approach to semantics in a first-order set-theoretic framework. Higher-order metaphysicians who think that semantics can only properly be done in higher-order languages might otherwise think that the challenge is self-undermining. Thanks to an anonymous referee for highlighting this point.

that our current theory is true and that we are defeasibly justified in believing it. The assumption that CH is determinately the case does nothing to block this possibility.

With this in mind, we can now see that the (putative) determinacy of higher-order claims does not help in sidestepping the challenge. Suppose, for instance, that HCH is perfectly determinate, that our preferred higher-order logic (which proves it) is true, and that we are defeasibly justified in believing it. Still, the epistemological challenge arises. We can still worry that we might easily have settled on a different logic — one that proves $H\neg CH$ — on the basis of similar abductive considerations (where logics are individuated syntactically).

Finally, it is far from clear that the language of higher-order logic is in fact determinate. The set-theoretic proofs purporting to show that it is are suspect if we doubt the determinacy of set theory itself. For instance, if we interpret the second-order quantifier via the powerset of the first-order domain, then the determinacy of the second-order quantifier depends on the determinacy of the powerset operation — and thus on the determinacy of set theory. Nor are higher-order proofs of the determinacy of higher-order language obviously better off. Much depends on our background metaphysics. Very generally, we seem to face a choice between two pictures of higher-order reality. On the first, we accept a specific higher-order logic but no higher-order meta-logics beyond it—this is the approach implicit in Tarski [1933]. On the second, we accept an infinite hierarchy of stronger and stronger higher-order logics, as in Tarski [1956], Rayo [2006], and Rayo and Linnebo [2012].[34]

In the first case, we cannot prove the determinacy of our logical language, since any such proof would require a higher-order meta-theory that we do not admit. So it is unclear what grounds we have for thinking that the logic is determinate. In the second case, we still lack any proof that *every* logic in the hierarchy is determinate, since such a proof would require a meta-theory stronger than any logic in the hierarchy itself. Thus, once again, it is unclear what grounds we have for believing in the determinacy of higher-order logic.[35]

4.9 Propositional Attitudes: Objection

It was suggested in the previous section that the only ingredient needed to get the epistemological worry going is the coherence of supposing that we could

[34]Thanks to an anonymous referee for emphasizing this point.

[35]This suggests that we have a vicious infinite regress with respect to the proofs of determinacy in the latter case. Consider the logic at the bottom level of the hierarchy, L_1. There is a proof of the determinacy of the language at L_1 at a higher-level. But this is not much of a proof if the language of that higher level is not determinate. Of course, there is a proof of the determinacy of the language of this higher level at a yet higher level. But again, this is not much of a proof if this language in turn is not determinate. And so on. In order for the proof at the very first level to work – to be convincing – the proof at every level has to work – the terms at every level have to have unique interpretations. But this is precisely what we do not have. If you are already suspicious of the determinacy of higher-order logic, you are unlikely to be swayed by this line. In general, it seems as though the proofs of determinacy bleed away up the hierarchy. From this perspective, they just beg the question.

easily have come to believe a different higher-order logic (individuated syntactically), using the same method we actually used to form our current beliefs. But some have argued that this supposition is not coherent.

There is a growing consensus among higher-order metaphysicians in favor of *intensionalism*: the view that absolute necessary equivalence (i.e., Top-identity) is sufficient for propositional identity. On this view, all axioms of the logic, qua propositions, are identical to *Top*, and so are the claims that these axioms — understood as sentences — express *Top*. If epistemic notions such as knowledge and safe belief are relations to propositions, then surely, if we safely believe anything, we safely believe *Top*. But then we safely believe all true logical axioms and all true claims about their meaning. It becomes difficult, even incoherent, to entertain the possibility that we could have accepted a logic inconsistent with the one we now accept, since this would require safely believing the bottom proposition.

Furthermore, if intensionalism is correct, then standard epistemological notions like knowledge, safe belief, and acceptance — understood as propositional attitudes — may do little or no explanatory work in the epistemology of logic. This raises a challenge for the dialectical force of the companions-in-guilt strategy: if the Benacerraf challenge is framed in terms of propositional attitudes, and those attitudes are unavailable or ill-defined within the intensionalist framework, then higher-order metaphysicians can dismiss the analogy with set theory. Even if epistemological worries arise in set theory — where belief and acceptance are typically construed as relations to propositions — these worries do not straightforwardly transfer to higher-order logic. From the standpoint of what appears to be the dominant view about the grain of propositions in higher-order metaphysics, epistemology framed in terms of propositional attitudes is simply not in good standing, and so cannot ground challenges that should trouble higher-order metaphysicians.[36]

4.10 Propositional Attitudes: Response

These are important objections, and I think they deserve careful attention. My response turns on the idea that everyone, including intentionalist higher-order metaphysicians, must rely on notions like knowledge and belief that are fine-grained enough to support a workable epistemology of logic. Perhaps this involves treating belief as a relation to syntactic objects such as sentences. I need not commit to a particular theory here. The key point is simply that the Benacerraf challenge for higher-order logic can be recast in terms of whatever fine-grained account of belief and justification higher-order metaphysicians accept.

It may not seem remotely plausible to such metaphysicians that they could easily have accepted a *different* higher-order logic if this is understood to mean accepting the bottom proposition. But it is not so implausible to suppose that they could easily have accepted a *different collection of logical axioms*, under-

[36]Thanks to an anonymous referee for raising this objection.

stood as sentences. Even if they are wary of epistemological challenges framed in terms of propositional attitudes, it is not clear why they should be similarly suspicious of challenges framed in terms of attitudes to sentences.

Moreover, nothing in the original Benacerraf challenge to set theory commits us to understanding safe belief or acceptance as a relation to propositions. We can just as well treat acceptance and safe belief as relations between agents and collections of sentences. Indeed, Field — the primary architect of the canonical version of the Benacerraf problem — is both an anti-realist about propositions and a deflationist about truth. He presumably could not have had propositional attitudes in mind.

So if the challenge has bite for set theory even when beliefs are construed as relations to sentences, then it should also have bite for higher-order logic on the same construal. As in the companions-in-guilt section: if there is no principled way to adjudicate between CH and ¬CH (understood as sentences), because each belongs to a theory that performs equally well under the abductive method, then we face a structurally identical problem for HCH and $H\neg CH$. So long as we individuate logical commitments syntactically — as we arguably must for purposes of logical epistemology — the challenge generalizes.

4.11 Humility: Objection

Let us now consider a final potential objection. Perhaps there is a Benacerraf problem for set theorists who endorse controversial claims, such as the Continuum Hypothesis, Martin's Axiom, or Large Cardinal axioms. But even if we grant this, it does not establish the companions-in-guilt thesis. For higher-order metaphysicians will likely never affirm, nor deny, higher-order analogues of such claims. In particular, they will likely never affirm either HCH or $H\neg CH$.

This reflects a broader methodological norm within higher-order metaphysics: namely, to adopt only as much logical strength as is required to formulate our best scientific theories. There is reason to believe that the strongest logic needed for this purpose is something like LF. Provided metaphysicians remain within the bounds of what is scientifically necessary, their higher-order beliefs remain epistemically safe, regardless of what set theorists happen to endorse (even if such safety would be lost by extending their commitments to stronger logics whose power outstrips scientific needs).

Thus, higher-order metaphysicians avoid the epistemological challenge simply by exhibiting appropriate epistemic humility. They refrain from taking controversial stances on higher-order analogues of set-theoretic claims, and thereby steer clear of the very conditions that generate the reliability worry.

4.12 Humility: Response

The crucial thing to note here is that nothing in the way I made the case peculiarly depends on the choice of CH, or indeed on any particularly recherche set-theoretic claim. I could just as well have made the prima facie case using intuitively safer set-theoretic claims, such as the axiom of choice, the axiom of infinity, or the axiom of powersets, higher-order versions of which are proved by currently popular logics including LF. Indeed, I could have argued that if there is a Benacerraf problem for set theorists who accept classical as opposed to constructive set theories, then there is also a problem for metaphysicians who accept classical as opposed to constructive higher-order logics. And while you might think making the case using these claims lessens its dialectical force, because there may be less reason to think that there is a Benacerraf problem for them, it in no way undermines the companions in guilt point. I used CH because it is the canonical example of an unsafe claim in set theory, not because it stacks the deck in favor of the argument.

Perhaps it is interesting in this context to note, that while the majority of set theorists' concerns have centered on claims such as CH, they have also worried about choice, replacement, the empty set, foundation, infinity, even bits and pieces of the underlying classical first-order logic such as double-negation elimination, and so on. If higher-order metaphysicians only ever committed themselves to LF, this would not shield them from these worries about claims such as these being transposed.

Moreover, this line of objection does not undermine the claim that if there is a Benacerraf problem for set theory regarding claims such as CH, there is also one for higher-order logic regarding claims such as HCH. It just says that we can mitigate the consequences of this – that we can avoid having unsafe logical beliefs – by neither accepting nor rejecting strong logical claims. But this seems like a fraught position to occupy as a higher-order metaphysician. Are higher-order metaphysicians supposed to ignore the question of whether HCH is true? On what grounds can they legitimately say that they do not need to grapple with this? After all, the game seems to be to try to land on the logic that allows us to prove everything that needs to be proved in all of science. But is not higher-order logic itself part of science? If we think it is incumbent on set theorists to consider CH, surely it is incumbent on higher-order metaphysicians to consider HCH.

Indeed, in this respect higher-order metaphysicians seem worse off than set theorists. Set theorists, by their own lights, may be able to fall back on the idea that CH is indeterminate. But this does not seem to be available to higher-order metaphysicians, by their own lights. There is a very interesting theorem of LF. Goodsell and Yli-Vakkuri [MS] call it the necessity of logic. It says that for every claim p in the language of pure higher-order logic, $\Box p \vee \Box\neg p$. Every purely logical claim is either identical to Top or identical to Bottom. We might think that this commits the higher-order metaphysician to grappling with any claim in the language of pure higher-order logic. Indeed, given this, we might think that accepting LF, while refusing to countenance stronger extensions which may settle HCH because these go beyond what we could abductively justify accepting, *is simply to admit* that there is a Benacerraf problem for logical claims which are independent of LF.

But even if you disagree, thinking that restricting ourselves to comparatively weak higher-order claims dissolves the problem, the point still remains that this move is equally available to the set theorist. So again, I emphasize that the point I am making in this paper is that set theory and higher-order logic are in the same boat with respect to the Benacerraf challenge, not that the challenge cannot be met in either case.

5 Pluralism and Safety

The upshot, if set theory and higher-order logic are indeed companions in guilt when it comes to the Benacerraf challenge, is that we should be equally concerned by the challenge in both cases. What this amounts to, of course, depends on one's view of whether the challenge can be met by set theorists, and whether the apparent lack of safety is something to be particularly worried about. In this section, I explore one prominent way in which some philosophers of mathematics and set theorists — those who have taken the challenge most seriously — have responded. I then suggest that, if we feel similarly, we might consider transposing their response to the case of higher-order logic.

The Benacerraf challenge has generally been accepted as a serious problem for set theory on the standard realist picture.[37] As mentioned in the introduction, some mavericks have responded by fundamentally rethinking set-theoretic reality. They have adopted various forms of set-theoretic pluralism.[38] There are differences among these views, of course, but the core idea is that there is not just one universe of sets, but many.[39] This reconception offers a potential solution to the epistemic worry while opening up exciting avenues of research, such as multiverse set theory,[40] the modal logic of forcing,[41] and set-theoretic geology.[42] It significantly expands our conception of mathematical reality.

To see how going pluralist may help, consider the view outlined in Balaguer [1995]. According to Balaguer, every classically consistent set theory is true of

[37]See Benacerraf [1973], Field [1989], Warren [2017], and Clarke-Doane [2020(b)] for classic and current arguments that there is a Benacerraf problem for set theory on the standard picture.

[38]See Balaguer [1995], Linsky and Zalta [1995], Hamkins [2012], and Priest [2024] for some notable examples.

[39]The precise statement of these views is often very tricky. It may seem that adopting a meta-theory other than set theory, such as a very strong higher-order logic + syntax, would readily allow a precise statement. Roughly, we could say that there are lots of properties and relations, set_0, $member_0$; set_1, $member_1$; set_2, $member_2$, and so on, such that the axioms of each set theory of interest are true on an interpretation that assigns some such property to "set" and some such relation to "member". The problem with this, as an anonymous referee recognized, is that this seems to be just another way of saying that the strongest one is unequivocally true. This has also been recognized by set-theoretic pluralists. As such, they have opted for other approaches. Steel [2014] states his multiverse view in a two-sorted first-order language with variables for sets, and variables for universes. More common though, has been to stick with just first-order set theory as the meta-theory, and to accept that the pluralist view can be neither formalized, nor precisely stated. This is the approach that Balaguer [1995], Hamkins [2012], and Clarke-Doane [2020 (b)] adopt. While this saddles these views with a measure of meta-theoretic instability, and ineffability, it does ensure that the views are genuinely pluralist.

[40]See Hamkins [2012].

[41]See Hamkins and Löwe [2008].

[42]See Fuchs, Hamkins, and Reitz [2015].

some distinct set-theoretic universe. Two features of his view are worth highlighting. First, both $ZFC + CH$ and $ZFC + \neg CH$ are classically consistent (assuming ZFC is), and so both CH and $\neg CH$ are true — though not in the same universe. There is no contradiction, since these statements concern different kinds of sets. Second, Balaguer adopts a highly cooperative metasemantics: let S be the set theory one accepts. Then set-theoretic sentences, when uttered, are automatically about a universe in which S is true.

Suppose, then, that we believe the set theory $ZFC + V = L$. Let s be the sentence $'2^{\aleph_0} = \aleph_1'$. Further, suppose that $ZFC + V = L$ is consistent, and thus true of some set-theoretic universe. We concede that we could easily have accepted instead the sentence $'2^{\aleph_0} < \aleph_1'$, and accepted $ZFC + \neg CH$. The crucial point, and this is where the very cooperative metasemantics comes in, is that in that scenario, we would not have had a false belief. For we would not have had beliefs about the $ZFC + V = L$ universe in which CH is true. We would have had beliefs about the $ZFC + \neg CH$ universe.

In general, had we adopted any classically consistent set theory, our set-theoretic beliefs would have been about a corresponding universe in which that theory holds. Hence, we could not easily have had false (but consistent) set-theoretic beliefs. Only inconsistent theories pose a problem. But Balaguer claims we are reliable judges of classical consistency. If so, then the pluralist, while assuming that their beliefs are defeasibly justified and true, can plausibly account for the safety of their set-theoretic beliefs.

If we are troubled by the Benacerraf challenge for higher-order logic, one intriguing response is to adopt a form of higher-order pluralism. Prima facie, there is reason to think that suitably robust pluralist assumptions could secure the safety of our higher-order logical beliefs. Roughly, if we can (i) delineate a class of higher-order logics that are all true in the sense that each governs a distinct subject-matter, and (ii) argue that our abductive methodology could not easily have led us to endorse a logic outside that class, then, granting that our beliefs are true and defeasibly justified, we may be able to show that they are safe. As Beall once quipped: if the target is hard to hit, make the target bigger!![43]

Of course, such a view would have to differ significantly from Balaguer's. I am reasonably confident that Balaguer does not conceive of set theory as a foundational mathematical theory, nor as a theory to be applied completely generally across domains such as physics and metaphysics.[44] By contrast, the interpretation of higher-order logic that interests higher-order metaphysicians is one on which it serves as a general-purpose foundational metaphysical theory, applicable across all the sciences. A pluralist view of higher-order logic must therefore preserve this generality.

This requirement distinguishes our proposal from Balaguer's and other forms of pluralism in at least two significant respects. First, if we want the higher-order logics we endorse to play the role of foundational metaphysical theories,

[43]See Beall [1999] for an elaboration on this point.

[44]Indeed, most pluralist views of set theory appear to almost studiously avoid these ambitions. Hartry Field is a notable exception.

then the criteria for logical truth must be suitably tailored. The claim that every classically consistent higher-order logic is true, for example, will clearly not suffice.[45] Instead, I tentatively suggest that a higher-order logic is true just in case it allows us to prove everything we need to prove in all areas of science. That is, it is a logic that we could, in principle, settle on using the abductive method.[46]

The second respect, given that we want all of the higher-order logics we endorse to be generally applicable, concerns how the claims of apparently contradictory logics can generally apply to math, physics, and metaphysics. The claim that different higher-order logics are true of different abstract structures would pretty clearly give up on this crucial feature. And it is not at all obvious that maintaining both pluralism and general applicability does not lead pretty swiftly to contradiction. One approach which I think has some promise is to adopt a kind of higher-order quantifier variance.[47] Roughly put, quantifier variance holds that there are multiple unrestricted quantifiers, each equally metaphysically privileged (whatever exactly that means), but which yield different answers to basic quantificational questions. For example, the unrestricted quantifier of the mereological nihilist implies that no composite objects exist,[48] while that of the mereological vitalist permits composition only when the parts constitute a life.[49] On this view, these quantifiers belong to different languages, all of which are equally good at carving nature at its joints —assuming any of them do.

Higher-order quantifier variance may offer us exactly what we need. On such a view, there would be a variety of different higher-order languages which

[45] For instance, there are classically consistent logics that are inconsistent with basic and seemingly uncontroversial physical and metaphysical claims. One such logic posits that there are precisely two individuals. Others are simply too weak to serve as foundations for general scientific or metaphysical theorizing.

[46] A fully developed account would need to clarify what it means for a logic to be, in principle, one we could settle on using the abductive method. This would involve a detailed analysis and elaboration of the relevant idealizations — perhaps drawing on a modified version of the classic pragmatist idea of the theories we would accept at the end of inquiry. Rather than invoking a single, ideal theory, however, the pluralist framework I am working with suggests appealing to the range of logics we could accept in ideal epistemic conditions.

The condition, as stated, is not yet informative with respect to the specific criteria such logics would satisfy; identifying and justifying such criteria is likewise a task for future work. That said, I expect the following may be among them: (i) the logic must be classically consistent; (ii) it must be consistent with some kernel of basic truths — e.g., basic physical or metaphysical principles; (iii) it must be able to serve as a foundation for classical mathematics (excluding set theory) in a manner similar to ZFC; and (iv) it must at least be able to interpret the modal logic KT.

While it is unlikely to be this simple, ideally the fully worked-out proposal might take something like the following form: let T be the theory incorporating the physical and metaphysical theories we would accept at the end of inquiry (perhaps some version of string theory together with a complete theory of the essences of the fundamental physical entities); then the acceptable logics might be precisely the consistent extensions of $LF + T$. Future work would also need to explore the justificatory status of such a constraint and whether further logical, conceptual, or practical conditions ought to be included.

[47] This idea is suggested and discussed in Clarke-Doane and McCarthy [2022].

[48] See Sider [2013].

[49] See Van Inwagen [1990].

are equally good with respect to the purpose of investigating the foundations of logic, math and metaphysics. Which is to say that they all cut at the joints if any do (whatever exactly we take this mean). In these different languages, different higher-order logics would be true. The idea being that these languages are precisely the ones in which there is a true higher-order logic which we could ultimately settle on using the abductive method. And that as such, for each higher-order logic which we could ultimately settle on using our abductive method, there is a language, which is a member of a class of equally metaphysically perspicuous languages, in which it is true. For instance, there may be a language L_{CH}, in which $LF + HCH$ is true, and a language $L_{\neg CH}$ in which $LF + H\neg CH$ is true.[50]

I tentatively propose the following view:

> For precisely all and only the logics that we could ultimately settle on using our abductive method, there is a metaphysically perspicuous language in which it it is true.

Much more needs to be said about the details and motivations of this pluralist approach to higher-order logic. Whether it can truly meet the Benacerraf challenge remains to be seen. But I hope to have made a plausible case that such a view is coherent, metaphysically serious, and worth pursuing — both in its own right and as a promising route through the epistemological thickets that threaten higher-order logical realism.[51]

[50] If such logics are ones we could ultimately settle on.

[51] I will mention a few things about one detail here, as I imagine it will occur to many readers. Embracing quantifier variance allows us to hold that there are different unrestricted quantifiers at every level, and thus to seemingly secure that apparently mutually contradictory completely general logics can all be true (in different languages). But this comes with a price. It amounts accepting that there is not an objective higher-order metatheory, in which to state the view, and in which to realize all of the acceptable higher-order logics at once. To be clear, this does not mean that we cannot assert the existence of intended interpretations of higher-order logics we like, one by one as it were, by adopting different higher-order metatheories. It just means that we cannot assert that they exist all at once. As such, this view is very different to a kind of view which an anonymous referee informed us is accepted by higher-order metaphysicians who like systems that are strong enough for semantics. In those systems, and in the metatheories (of transfinite orders) that they can be extended to for finite-order higher-order logics, it can be proved that there is an interpretation on which all the theorems of any well-known system of higher-order logic are all true. But this involves accepting a very strong logic, and a metatheory in which we can prove that the favored strong system (or the part of it that can be formulated in the object language) is the only one that's about $\forall^\sigma$. Which is to say it involves accepting that there is a uniquely good interpretation of the higher-order vocabulary.

While this may be unpalatable to some higher-order metaphysicians, I am mostly untroubled. Pluralists have to accept that they cannot assert all at once the existence of intended interpretations of all of the logics they accept. But I think there is a good case to be made that higher-order metaphysicians who accept standard, non-pluralist views also have difficulty doing this. As I outlined in section 4.8, either we accept a specific higher-order logic and no further higher-order meta-logics, or we accept an infinite hierarchy of stronger and stronger higher-order logics. In the first case we cannot assert the existence of an interpretation of our logic, as to do so would require using a meta-logic of a higher-order. And in the second case, while we can use logics in the hierarchy to assert the existence of interpretations of

6 Conclusion

Higher-order metaphysicians are currently engaged in a fascinating and potentially highly fruitful foundational project. However, as I have outlined in this paper, higher-order logic appears to be just as susceptible to the Benacerraf challenge as set theory. As such, one should be equally troubled in both cases, even if one strongly prefers higher-order logic to set theory on abductive grounds.

I have suggested that if the challenge is troubling, a potentially promising response would be to adopt a pluralist approach to higher-order logic. Higher-order pluralism may offer an epistemologically safer foundation for metaphysical theorizing. Of course, this would be a very different kind of foundation than what higher-order metaphysicians currently seem to seek. Rather than positing a single, privileged higher-order logic within which to theorize, pluralism invites us to recognize a variety of higher-order logics and to select whichever is most apt for the theoretical purpose at hand.

The consequences of adopting such a view are difficult to overstate. Embracing higher-order logical pluralism would amount to a fairly radical rethinking of the nature of validity, modality, and metaphysics more broadly, much as set-theoretic pluralism has already prompted a radical rethinking of the foundations of mathematics. And just as in the case of set theory, I do not believe this would be a negative result. Far from undermining the project of higher-order metaphysics, such a shift might open up new avenues of inquiry and offer a more flexible, resilient foundation for the field.

7 Bibliography

- Arntenius, F. & Dorr, C. (2012). Calculus as Geometry, In F. Arntzenius (ed.), *Space, Time and Stuff.* Oxford: Oxford University Press.
- Bacon, A. (2018). The Broadest Necessity. *Journal of Philosophical Logic*, 47 (5):733-783.
- — & Dorr, C. (2024). Classicism. In P. Fritz & N. K. Jones (Eds.), *Higher-order Metaphysics* (pp. 109–190). Oxford: Oxford University Press.
- Balaguer, M. (1995). A Platonist Epistemology. *Synthese*, 103(3), 303–325.
- Beall, J. C. (1999). From Full Blooded Platonism to Really Full Blooded Platonism. *Philosophia Mathematica*, 7(3), 322–325.
- Benacerraf, P. (1973). Mathematical Truth. *Journal of Philosophy*, 70(19), 661–679.

logics which are lower in the hierarchy, it does not seem that we can assert all at once the existence of interpretations of all of the logics we accept. To do so, we would need to be able to talk about all the levels at once. As such, we might think that the non-pluralist and the pluralist are in the same boat when it comes to being able to assert all at once the existence of interpretations of all of the logics they they respectively endorse. Inexpressibility will rear its head for any semantically open foundational theory. It's (almost) everyone's problem.

- Chalmers, D. (2020). Debunking Arguments for Illusionism About Consciousness. *Journal of Consciousness Studies*, 27(5–6), 258–281.
- Church, A. (1940). A Formulation of the Simple Theory of Types. *Journal of Symbolic Logic*, 5(2), 56–68.
- Clarke-Doane, J. (2017). What is the Benacerraf Problem? In F. Pataut (Ed.), *New Perspectives on the Philosophy of Paul Benacerraf: Truth, Objects, Infinity*. Dordrecht: Springer.
- — (2020a). Set-theoretic Pluralism and the Benacerraf Problem. *Philosophical Studies*, 177(7), 2013–2030. doi:10.1007/s11098-019-01365-w
- — (2020b). *Morality and Mathematics*. Oxford: Oxford University Press.
- — & McCarthy, W. (2022). Modal Pluralism and Higher-Order Logic. *Philosophical Perspectives*, 36(1), 31–58.
- Field, H. (1980). *Science Without Numbers: A Defence of Nominalism*. Princeton: Princeton University Press.
- — (1989). *Realism, Mathematics and Modality*. Oxford: Basil Blackwell.
- — (1994). Are Our Mathematical and Logical Concepts Highly Indeterminate? *Midwest Studies in Philosophy*, 19(1), 391–429.
- Florio, S. & Incurvati, L. (2019). Metalogic and the Overgeneration Argument. *Mind*, 128(511), 761–793. doi:10.1093/mind/fzy042
- Fuchs, G., Hamkins, J. D. & Reitz, J. (2015). Set-theoretic Geology. *Annals of Pure and Applied Logic*, 166(4), 464–501.
- Goodsell, Z. & Yli-Vakkuri, J. (MS). LF: A Foundational Higher-Order Logic. Manuscript.
- Hamkins, J. D. (2012). The Set-Theoretic Multiverse. *Review of Symbolic Logic*, 5(3), 416–449.
- — & Löwe, B. (2008). The Modal Logic of Forcing. *Transactions of the American Mathematical Society*, 360, 1793–1817.
- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lewis, D. (1984). Putnam's Paradox. *Australasian Journal of Philosophy*, 62(3), 221–236.
- Linnebo, Ø. & Rayo, A. (2012). Hierarchies Ontological and Ideological. *Mind*, 121(482), 269–308.
- Linsky, B. & Zalta, E. (1995). Naturalized Platonism versus Platonized Naturalism. *The Journal of Philosophy*, 92(10), 525–555.

- Priest, G. (2024). *Mathematical Pluralism.* Cambridge: Cambridge University Press.
- Putnam, H. (1980). Models and Reality. *Journal of Symbolic Logic*, 45(3), 464–482.
- Quine, W. V. O. (1948). On What There Is. *Review of Metaphysics*, 2(5), 21–38.
- — (1951). Two Dogmas of Empiricism. *Philosophical Review*, 60(1), 20–43.
- Rayo, A. (2006). Beyond Plurals. In A. Rayo & G. Uzquiano (Eds.), *Absolute Generality.* Oxford: Oxford University Press.
- Schechter, J. (2010). The Reliability Challenge and the Epistemology of Logic. *Philosophical Perspectives*, 24(1), 437–464.
- Shapiro, S. (1991). *Foundations Without Foundationalism: A Case for Second-Order Logic.* Oxford: Oxford University Press.
- Sider, T. (2013). Against Parthood. *Oxford Studies in Metaphysics*, 8, 237–293.
- Steel, J. (2014). Gödel's Program. In J. Kennedy (Ed.), *Interpreting Gödel: Critical Essays* (pp. 153–179). Cambridge: Cambridge University Press.
- Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies*, 127(1), 109–166.
- Tarski, A. (1933). The Concept of Truth in the Languages of the Deductive Sciences. *Prace Towarzystwa Naukowego Warszawskiego*, 34, 13–172. English translation in Tarski (1956), pp. 152–278.
- — (1943). The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research*, 4(3), 341–376.
- — (1956). *Logic, Semantics, Metamathematics.* Oxford: Clarendon Press.
- Trueman, R. (2020). *Properties and Propositions: The Metaphysics of Higher-Order Logic.* Cambridge: Cambridge University Press.
- van Inwagen, P. (1990). *Material Beings.* Ithaca: Cornell University Press.
- Warren, J. (2017). Epistemology Versus Non-Causal Realism. *Synthese*, 194(5), 1–26.
- Williamson, T. (2013). *Modal Logic as Metaphysics.* Oxford: Oxford University Press.