Justin Clarke-Doane

Columbia University

[This is a penultimate draft of a paper that is forthcoming in *Philosophical Studies.*]

Set-Theoretic Pluralism and the Benacerraf Problem[1]

Set-theoretic pluralism is an increasingly influential position in the philosophy of set theory (Balaguer [1998], Linksy and Zalta [1995], Hamkins [2012]). According to the pluralist, "whenever you have a consistent [formulation of set theory], then there are…objects that satisfy that theory *under a perfectly standard satisfaction relation*…[A]ll the consistent concepts of set …are instantiated side by side [Field 2001, 333, emphasis in original]." Of course, the Completeness Theorem ensures that every consistent theory – set-theoretic or otherwise – has a model. What the pluralist adds is that it has an *intended* model.[2] The intuition is that set-theoretic "truth comes cheaply" (given consistency), but not because it depends on us. Set-theoretic truth comes cheaply because the set-theoretic universe – or, better, pluriverse – is so rich, and the semantics of set-theoretic discourse so cooperative, that consistent theories are automatically about the entities of which they are true, and there are always such entities.

There is considerable room for debate about how best to formulate set-theoretic pluralism, and even about whether the view is coherent.[3] But there is widespread agreement as to what there is

[1] Thanks to Joel David Hamkins, Achille Varzi, Jared Warren, and audience members of the *Set-Theoretic Pluralism: Indeterminacy and Foundations* conference at the University of Aberdeen for helpful discussion.

[2] The "intended model" is not strictly speaking a model at all, since it is not a set.

[3] For more on this, see Section 2.

to recommend the view (given that it can be formulated coherently). Unlike set-theoretic "universalism", set-theoretic pluralism affords an answer to Benacerraf's epistemological challenge.[4] By set-theoretic "universalism", I mean the view associated with Godel which "takes as basic some one conception of set, and constructs out of sets so conceived all other mathematical objects [Field 2001, 333]." The problem for the universalist is to explain how it is that we happened to land on the axioms true of the "one true V".[5] The pluralist, by contrast, is supposed to face no such problem. The following quotations are representative.

> The most important advantage that [pluralism] has over [non-pluralist] versions of platonism...is that all the latter fall prey to Benacerraf's epistemological argument [Balaguer 1995, 317].

> [Pluralist views] allow for...knowledge in mathematics, and unlike more standard platonist views, they seem to give an intelligible explanation of it [Field 2005, 78].

[4] Although this is the standard argument for set-theoretic pluralism, Hamkins [2012] suggests that the view also does better justice to set theorists' experience working with different models of set theory. He writes,

> This abundance of set-theoretic possibilities poses a serious difficulty for the universe view...one must explain or explain away as imaginary all of the alternative universes that set theorists seem to have constructed. This seems a difficult task, for we have a robust experience in those worlds….The multiverse view...explains this experience by embracing them as real [2012, 418].

But it is unclear what this argument comes to. Surely the real existence of the models in question does not help to causally explain set theorists' psychological states. For more on this, see Section 2.

[5] This formulation makes it sound as if the universalist accepts the Axiom of Foundation, which should not be built into the view. But the terminology is entrenched, so I stick to it here.

> [The pluralist has] an answer to Benacerraf's worry that no link between our cognitive faculties and abstract objects accounts for our knowledge of the latter [Linsky and Zalta 1995, 25].

> [The pluralist's] strong existence assumptions imply that, so long as we know that our full conception characterizing a theory is consistent, it can't fail to be true of…[a] portion of mathematical reality, so that knowledge of mathematical truths is reduced to knowledge of the consequences of consistent mathematical theories [Leng 2009, 124].

> [Pluralism]...solve[s] the problem by expanding platonic heaven to such a degree that one's cognitive faculties can't miss it (as it were). (If you're having trouble hitting the target, then just make your target bigger!...) [Beall 1999, 323].

The purpose of this paper is to determine what Benacerraf's challenge could be such that these views are warranted. I argue that it could not be any of the challenges with which it has been traditionally identified by its advocates, like of Benacerraf and Field. Not only are none of the challenges easier for the pluralist to meet. None satisfies a key constraint that has been placed on Benacerraf's challenge, independent of the universalism-pluralism debate. However, I argue that Benacerraf's challenge could be the challenge to show that our set-theoretic beliefs are *safe* – i.e., to show that we could not have easily had false ones (using the method that we actually used to form ours). Whether the pluralist is better positioned to show that our set-theoretic beliefs are

safe turns on a broadly empirical conjecture which is outstanding. If this conjecture proves to be false, then it is unclear what the epistemological argument for set-theoretic pluralism could be.

1. The Reliability Challenge

In "Mathematical Truth", Benacerraf writes,

> [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….We accept as knowledge only those beliefs which we can appropriately relate to our cognitive faculties [1973, 674].

The request for an "account of the link between our cognitive faculties and the [mathematical] objects known" can be interpreted in at least two ways. First, it can be interpreted as the challenge to explain the (defeasible) *justification* of our set-theoretic beliefs. Second, it can be interpreted as the challenge to explain the *reliability* of those beliefs. (It can also be understood as the challenge to *justify*, in the dialectical sense, our mathematical beliefs.) The second challenge is widely taken to be the more serious of the two (Enoch [2010], Clarke-Doane [2016a]). To see why, consider Godel's suggestion that "despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as true [1947, 483-4]." Godel complains,

"I don't see any reason why we should have less confidence in this kind of perception, i.e., mathematical intuition, than in sense perception…[1947, 483-4]" Godel's remarks are widely regarded as being unresponsive to the most pressing aspect of Benacerraf's epistemological challenge. The reason is that even if appeals to "intuition" help to explain the (defeasible) justification of our set-theoretic beliefs – help to explain, that is, why it would be reasonable to trust their contents, absent reason to doubt them – they do nothing to explain their reliability. They leave it mysterious why being the content of an "intuition" would be a reliable symptom of being true.[6]

Although Benacerraf's first drew general attention to the problem, the canonical formulation of the epistemological challenge is due to Field (Liggins [2010], Linnebo [2006]). He writes,

> We start out by assuming the existence of mathematical entities that obey the standard mathematical theories; we grant also that there may be positive reasons for believing in

[6] A similar problem plagues indispensability-based epistemologies inspired by Quine. Colyvan writes,

> [L]et's take a…charitable reading of the… challenge…to explain the reliability of our systems of beliefs….[W]e see that Quine has already answered it: we justify our system of beliefs by testing it against bodies of empirical evidence Colyvan [2007, 111].

Even if we can explain the justification of (or justify, in a dialectical sense) our belief in sets in in the way that we can explain the justification of (or justify) our belief in electrons, it does not follow that we can explain the reliability of our belief in sets in the way that we can explain the reliability of our belief in electrons. As Leng writes,

> Mathematical objects are...acausal and non-spatiotemporal….These...features put them on a...different footing than electrons…[W]e should expect that the observed phenomena would be very different on the hypothesis that there are no such things [as electrons]...But if such counterfactual considerations have force against those sceptical about the unobservable physical objects posited by our theories, no analogous counterfactual is available against those sceptical about the mathematical objects our theories posit. A mathematical realist who starts a challenge, 'If there *were* no numbers, then...' will find it difficult to finish this supposed counterfactual in a way that could trouble those sceptical of mathematical objects [2010, 202, italics in original].

> those entities… But Benacerraf's challenge … is to … explain how our beliefs about these remote entities can so well reflect the facts about them … *[I]f it appears in principle impossible to explain this*, then that tends to *undermine* the belief in mathematical entities, *despite* whatever reason we might have for believing in them [Field 1989, 26, emphasis in original].

Field's formulation of Benacerraf's challenge has a number of virtues. First, it cannot be dismissed as a puzzle of no practical significance. The apparent impossibility of answering the challenge is supposed to *undermine* our set-theoretic beliefs.[7] If it appears impossible to answer, then we ought to *change our set-theoretic beliefs*. Note that the challenge would not obviously have this significance if it merely purported to show that our set-theoretic beliefs fail to qualify as knowledge. If our set-theoretic beliefs were justified and we could "explain their reliability" then it is hard to see why we should give them up *even if* they failed to qualify as knowledge.[8]

Second, Field's is not a "convince the skeptic" challenge. Field grants both the (actual) truth and (defeasible) justification of our set-theoretic beliefs. His contention is that it appears impossible

[7] Assuming a "truth-value realist" interpretation, that is. For more on the distinction between "ontological" and truth-value realism, see Shapiro [2000], and the fourth point below. (I will speak of "our" set-theoretic beliefs in what follows to refer to those of the set-theoretic community – bracketing the familiar complication that it is unclear how to understand the notion of a scientific community's beliefs. Field does not claim that we must explain the reliability of every individual's set-theoretic beliefs, however outlandish, in order to answer his challenge.)

[8] It is sometimes suggested that Field's formulation of Benacerraf's challenge would not undermine our set-theoretic beliefs, but would merely undermine a particular construal of their contents ("truth-value realism"). But this is doubtful – at least assuming that truth-value realism was plausible to start with. In general, when faced with undermining – or, indeed, rebutting – evidence we do not simply reinterpret the contents of our beliefs. We give them up. For example, suppose we use a machine to test a liquid for a compound. We then learn that it only gave a positive reading because it was stuck on "positive". Then our belief that the compound is in the liquid, not our belief that facts about it do not depend on our beliefs, seems undermined. *Perhaps* matters are different when a whole class of beliefs is on trial. But, if so, some explanation of the difference is needed. In any event, nothing non-semantic turns on whether we take the unanswerability of the reliability challenge to undermine our mathematical beliefs or merely to undermine our belief in mathematical realism.

"explain their reliability" *even given these assumptions*. If Field did not grant these things, then Benacerraf's challenge would overgeneralize. After all, the evolutionary explanations of our having reliable mechanisms for perceptual belief, and the neurophysical explanations of how those mechanisms work such that they are reliable, all presuppose the reliability of our perceptual beliefs (even if they do not state that they are reliable) [Schechter 2010]. Field's contention is that there is a relevant *difference* between our set-theoretic beliefs and our perceptual beliefs.

Third, Field's challenge is distinct from the challenge to explain the *determinacy* of our set-theoretic beliefs (Putnam [1980], Field [1989, Introduction]).[9] Of course, skepticism about the reliability of our set-theoretic beliefs might engender skepticism about their determinacy, especially if that skepticism stems from causal considerations. But Benacerraf's challenge is supposed to arise *even if* "is a member of" has a determinate extension (on an occasion of use).

Finally, despite Field's talk of "objects", his formulation of the challenge does depend on an ontologically committal interpretation of set theory.[10] What matters is that set-theoretic truths would be causally, counterfactually, and constitutively independent of human minds and languages. Indeed, if this were not so, then Benacerraf's challenge would have no analog for moral realism in the context of nominalism about universals, or for realism about "metaphysical" possibility, so long as modal operators were taken as primitive. But this is highly

[9] Barton [2016] seems to conflate these challenges in his discussion of pluralism.
[10] See Hellman [1989] or Chihara [1990] for ontologically-innocent interpretations of our mathematical theories. It is no wonder that Field takes the problem to depend on an ontologically committal interpretation of mathematics since he himself appeals to primitive modal ideology. See, again, his [1989, Introduction].

counterintuitive. Even if the mind-and-language independent truth of such sentences as "giving to charity is morally obligatory" or "there could have been a 1000-story skyscraper" do not owe anything to the existence of The Obligatory or a world with a 1000-story skyscraper, there is surely a question of what explains the reliability of our belief in such sentences (given that they are mind-and-language independently true).

2. The Standard View

It is widely agreed that, whatever the costs of set-theoretic pluralism (and supposing that the view is coherent), the principal consideration favoring the view is that it affords an answer to Benacerraf's challenge (Balaguer [1998], Beall [1999], Linsky and Zalta [1995], and Field [2005]). Indeed, pluralism is widely taken to be the only view consistent with platonism that does this (though, again, it need not be formulated as a platonistic view) (Balaguer [1998, 51], Field [2005, 78]). The reason is simple. The set-theoretic pluralist holds that the set-theoretic pluriverse is sufficiently rich, and that the semantics of set-theoretic discourse sufficiently cooperative, that *all* consistent formulations of set theory are true of their intended subjects – not just true of a model. Consequently, all we have to do in order to have true set-theoretic beliefs is to have consistent ones. The gap between consistency and truth dissolves. By contrast, on the traditional view, associated with Godel [1947] – which I will call "universalism" – the overwhelming majority of consistent formulations of set theory are false, just as the overwhelming majority of consistent formulations of a physical theory are false. If we know that ZF is consistent, then we know – by theorems of Godel and Cohen – that so are ZF + CH and ZF

+ ~CH. For the pluralist, this is all there is to know about CH.[11] But, for a the universalist, a deep question remains. Is CH *true*?

Note that pluralism only ensures that "truth comes cheaply" in tandem with the "cooperative" semantics.[12] It must be the case that had we had different, but still consistent, set-theoretic beliefs, then we would have *changed the subject*. After all, even if the pluriverse contains "shmets" – where shmets are just like ZFC-sets except that, say, they fail to satisfy the Axiom of Choice – it does not follow that ZF + ~AC is about them. Perhaps reference in science is highly parasitic, and historically set theorists quantified over ZFC sets with their use of "there is". Then while the pluriverse would be rich, this would be of no obvious use to us, epistemologically speaking. We would still run the same risk of going wrong in speculating about "all sets".

Before turning to the question of what Benacerraf's challenge could be such that the pluralist, but not the universalist, can answer it, I need to discuss an apparent inconsistency in the view (Koellner [2013]). The pluralist presumably wishes to claim that our belief in ZF (or some other set theory interpreting PA) is true. Suppose, then, that ZF is true, and, therefore, consistent. Godel's Second Incompleteness Theorem ensures that, if ZF is consistent, then so is ZF + ~Con(ZF). But, according to set-theoretic pluralism, every consistent set theory is true. Hence, ZF + ~Con(ZF) is not just consistent, but true. But does not "~Con(ZF)" say that ZF is

[11] More exactly, this is all there is to know that is of any mathematical interest. One could still ask the semantic question of what the extension of "is a member of" is on an occasion of use. Alternatively, one might ask what is "packed into" *our* concept of set. But such questions are really just about us and lack any mathematical interest. They put no constraints on what sets (or set-like entities) there are.

[12] This point does not seem to me to be sufficiently recognized in the literature. Balaguer [1998], for instance, seems to me not to be sufficiently sensitive to this.

inconsistent in "ZF + ~Con(ZF)"? If it does, then ZF is inconsistent, and so not true. This is a contradiction.

It is in keeping with her "cooperative semantics" for the pluralist to deny that "~Con(ZF)" does mean that ZF is inconsistent out of its advocate's mouth (Warren [2015]). After all, the pluralist says that "every nonempty set has a Choice function" means one thing in the mouth of a believer in Choice and another in the mouth of a believer in Determinacy. If the advocate of ZF + Con(ZF) and ZF + ~Con(ZF) both speak truly, then they must mean different things by "proof" (and by "finite sequence"). But such a view is radical. It is tantamount to pluralism about consistency itself (Field [1998]). If the pluralist wishes to show that it is *objectively* the case that we can "explain the reliability" of our set-theoretic beliefs, then she must accept a limited amount of arithmetic objectivity. Otherwise, it would not even be an objective fact that, e.g., ZF is true, because it is consistent, and so certainly not an objective fact that we can explain the reliability of our belief in ZF! A still quite radical position which would secure the objectivity of consistency statements would be that any *Pi_1 sound* theory is true of its intended subject. On this view, while both of, e.g., ZF + AC and ZF + ~AC are true of their intended subjects, if ZF is consistent, only one of ZF + Con(ZF) and ZF + ~Con(ZF) is. We will see that this view seems as capable of answering Benacerraf's challenge, properly conceived, as the more radical view does. Accordingly, for concreteness, I will henceforth take set-theoretic pluralism to be the view that all Pi_1 sound theories are true of their intended subjects. But my argument will not depend on this.

With these qualifications out of the way, what is the interpretation of "explain the reliability" such that all of (a) – (c) below are true?

(a) It appears impossible to explain the reliability of our set-theoretic beliefs, assuming universalism.

(b) It does not appear impossible to explain the reliability of our set-theoretic beliefs, assuming pluralism.

(c) If it appears impossible to explain the reliability of our set-theoretic beliefs, then this undermines those beliefs.

Despite the widespread view that the pluralist is better positioned to answer Benacerraf's challenge than the universalist, I am not aware of an attempt to answer this question. But absent such an interpretation, we cannot translate the slogan that truth comes cheaply for the pluralist into an argument that the pluralist is better positioned than the universalist to answer Benacerraf's challenge. Moreover, if the pluralist is not better positioned to answer that challenge, then it is not clear what the epistemological argument for pluralism could be.

In what follows, I will seek an interpretation of "explain the reliability" which satisfies (a) – (c).

3. Causation and Explanation

The most familiar interpretation "explain the reliability" is suggested by Benacerraf. He writes,

> I favour 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 of mathematical truth makes it difficult to see how mathematical knowledge is possible….[T]he connection between the truth conditions for statements of number theory and any relevant events connected with the people who are suppose to have mathematical knowledge cannot be made out" [1973, 671-673].

Let us interpret Benacerraf's suggestion as follows.

> *Answer 1 (Causation)*: In order to "explain the reliability" of our set-theoretic beliefs, it is necessary to establish, for any one of our (token) beliefs that P, that there obtains a *causal relation* between our belief that P and the subject matter of P.

Component (a) is surely plausible under this interpretation. Nevertheless, (b) false if (a) is true. If there is no causal relation between our set-theoretic beliefs and the "one true V", then there is no causal relation between our set-theoretic beliefs and the "pluriverse". Indeed, set-theoretic pluralists explicitly purport to answer Benacerraf's challenge *under the assumption* that "we are not capable of coming into any sort of contact" with sets (Balaguer [1998, 51]).[13] Moreover,

[13] See also Linksy and Zalta [1995, 25].

even if *Answer 1* satisfied (b), it would not satisfy (c) (Field [2005, 77]). The causal theory of knowledge is widely rejected for reasons that are independent of Benacerraf's challenge. But if it is implausible that *knowledge* that P requires a causal relation to obtain between our belief that P and P's subject matter, then it is even more implausible that mere justified belief that P requires this.

But even if Benacerraf's own interpretation of "explain the reliability" does not satisfy (a) – (c), there is a proposal in the neighborhood which is more promising. It is one thing to require that the subject matter of our beliefs helps to *cause* those beliefs, and it is another to require that their contents (or disquotational truth) help to *explain* those beliefs. If we do not believe the likes of the Axiom of Pairing "because" it is true, then would not the reliability of our beliefs in such axioms be "merely coincidental"? Such a suggestion is familiar from metaethics. Sayre-Mccord writes,

> The problem with moral theory is that moral principles… appear not to play a role in explaining our making the [judgments] we do. All the…work seems to be done by psychology, physiology, and physics [1988, 442].

Let us interpret Sayre-Mccord's suggestion, in the present context, as follows.

> *Answer* 2 *(Explanation)*: In order to "explain the reliability" of our set-theoretic beliefs, it is necessary to establish, for any one of them, that P, that P helps to explain, even if not cause, our belief that P.

Answer 2, again, seems to satisfy (a). To be sure, there is an "indispensability" argument that, for at least a fragment of ZF, T, the explanation of our belief in any of T's consequences will imply that consequence. As Steiner notes, "something is causally responsible for our belief, and there exists a theory… which can satisfactorily explain our belief in causal style. *This theory, like all others, will contain [that fragment]*" [1973, 61, emphasis in original]. But showing that the content of our belief is implied by its explanation is not yet to show that that content "explains" that (token) belief.[14] Consider any recondite logical truth, P, that you believe. Then P is trivially a consequence of any explanation of your belief that P – simply because P is a consequence of every explanation at all. But surely the fact that P need not thereby have had a role in explaining your belief that P. You may have come to believe that P by flipping a coin, or via lucky but erroneous computation, for example. Indeed, had you believed ~P, instead of P, P, and not ~P, *still* would have been implied by any explanation of your coming to believe it. Let us, therefore, grant (a) for the sake of argument.

Nevertheless, if (a) is true, (b) is false. If we are worried that we do not believe Pairing "because" it is true, it is no consolation to learn that it is true in part of the pluriverse, rather than

[14] Of course, if one dismisses hyperintentional notions like "explanation", in the present sense, as unintelligible, then *Answer 2* does not get off the ground.

in the universe. Analogous considerations tell directly against the following weakening of *Answer 2*.

> *Answer 3 (Indispensability)*: In order to "explain the reliability" of our set-theoretic beliefs, it is necessary to establish that, for any one of them, that P, that P is implied by the best explanation of our belief that P, even if not "in an explanatory way".

But, again, the fact that Pairing is true in the pluriverse, rather than the universe, is no evidence that it is implied by the best explanation of our belief in Pairing. So, if (a) is true, then (b) is false.[15]

And, yet, even if (a) and (b) were both true, it is doubtful that either *Answer 2* or *3* would satisfy (c). The problem is that it is hard to see how learning that the best explanation of our belief fails to imply its content could undermine it *absent some reason to think that whether our belief co-varies with the truth depends on whether the truth is so implied*. But even if we have a reason to suspect that beliefs in *contingent* truths co-vary with the truths only if their contents are implied by their best explanation, we do not have reason to think that beliefs in *metaphysically necessary* truths co-vary with the truths only when this is the case. In particular, even if the contents of our set-theoretic beliefs are not implied by their explanation, it may be that we could not have easily had false such beliefs. Imagine, for example, that our set-theoretic beliefs (*not* true set-theoretic beliefs *per se*) are evolutionarily "hard-wired" in us.[16] Then, if the set-theoretic

[15] Moreover, we do seem to be able to show that the contents of at least some of our set-theoretic beliefs are implied by their explanation for the reason alluded to by Steiner in the quotation above. So, (a) is questionable as well.
[16] For an argument that not even true arithmetic beliefs *per se* are hard-wired in us, see Clarke-Doane [2012].

truths could not have easily been different, our set-theoretic beliefs could not have easily been false – even if their contents are not implied by their explanation. (Of course, I do not advocate this hypothesis. My point is just that learning that the best explanation of our set-theoretic beliefs fails to imply their truth would give us no reason to doubt it.) Moreover, assuming a standard semantics for counterfactuals, it is also true – albeit vacuously – that had the set-theoretic truths been different, our set-theoretic beliefs would have been correspondingly different (if the set-theoretic truths are metaphysically necessary). And while such a counterfactual may not be *non-vacuously* true, *neither is pretty much any counterpossible* of the form "had the F-truths been different, our F-beliefs would have been correspondingly different". For example, it seems false, if not vacuously true, that had it been the case that the "bridge laws" which link subvenient properties to supervenient properties been different, our beliefs would have been correspondingly different.[17]

Nor does learning that the explanation of our beliefs in metaphysically necessary truths fails to imply their contents seem to give us reason to doubt that the probability that they are true is high. Whether it gives us reason to doubt that the *epistemic* probability that they are true is high is what is in question. So, let the probability in question be "objective". Then, for any mathematical truth, P, presumably Pr(P) = 1, if mathematical truths are metaphysically necessary. (We *could* require that Pr(P) = 1 only if P is necessary in an even stronger sense -- e.g., "conceptually" necessary. But, then it also would turn out to be objectively improbable that, e.g., atoms arranged chair-wise compose a chair.) Moreover, even if P is not implied by

[17] For much more on this, see my [2016, 2.4 & 2.6] as well as chapters 4 and 5 of my [Forthcoming].

any explanation of our belief that P, it may be that Pr(we believe that P) ≈ 1, because the probability of our having the mathematical beliefs that we have is high. But (P & we believe that P) implies (our belief that P is true). So, it may be that Pr(our belief that P is true) ≈ 1 as well – even if P is not implied by any explanation of our belief that P.

These considerations suggest that we should consider interpretations of "explain the reliability" in terms of modal conditions. Let us turn to some now.

4. Counterfactual Dependence

Field explicitly identifies "explaining the reliability" of our set-theoretic beliefs with establishing the counterfactual dependence between our beliefs and the truths. He writes,

> The Benacerraf problem…seems to arise from the thought that we would have had exactly the same mathematical…beliefs even if the mathematical…truths were different …and this undermines those beliefs" [2005, 81].

What does Field mean by "different"? He apparently means arbitrarily different. He is particularly concerned about the case in which there are no sets at all [2005, 80—81].

Let us, therefore, interpret Field's suggestion as follows.

> *Answer 4 (Counterfactual Persistence)*: In order to "explain the reliability" of our set-theoretic beliefs, it is necessary to establish, for any one of them, that P, that had the set-theoretic truths been arbitrarily different such that ~P, we would not still have believed that P.

Answer 4 must be complicated. Notoriously, modal conditions on knowledge and justified belief require relativization to methods of belief formation. That our set-theoretic beliefs would have failed to be correspondingly different had the set-theoretic truths been different only because the closest worlds in which the set-theoretic truths are different are worlds in which we decide what to believe by flipping a coin is not undermining. Let us, thus, reformulate *Answer 4* as follows.

> *Answer 4 (Counterfactual Persistence)*: In order to "explain the reliability" of our set-theoretic beliefs, it is necessary to establish, for any one of them, that P, that had the set-theoretic truths been arbitrarily different such that ~P, we would not still have believed that P (had we used the method that we actually used to determine whether P).

Again, all counterfactuals conditionalizing on the set-theoretic truths being different are vacuous on a standard semantics, assuming that the set-theoretic truths are metaphysically necessary. But if they are, then *Answer 4* does not get off the ground. In particular, (a) is false. Let us suppose, then, that a non-standard semantics of counterfactuals is correct according to which there is a serious question as to what set-theoretic beliefs we would have had, had the set-theoretic truths been different. Then (a) does seem to be true. Had the "one true V" failed to exist, or had it

been different, it seems that we still would have believed, e.g., Pairing. Perhaps one could protest that had the set-theoretic truths been different, the physical truths would have been different, and our set-theoretic beliefs would have reflected the difference – since set theory is indispensable to empirical science. But this argument is highly suspect (Field [1989: 18–20]), even bracketing the fact that it would seem hopeless in cases where we merely vary recondite axioms, such as large cardinal hypotheses. Why think that we would believe a set-theoretic proposition only if it is indispensable to empirical science? Let us, therefore, grant (a) for the sake of argument.

Answer 4 is still untenable. First, (b) is false if (a) is true. If our set-theoretic beliefs would have been the same had the "one true V" failed to exist, or had it been different, then something similar would surely be true of the "pluriverse".[18] Pluralists explicitly acknowledge that "[i]f there were never any such things as [sets], the physical world [and, so, our set-theoretic beliefs] would be…as it is right now" (Balaguer [1999, 113]). Second, even if (b) were true, (c) is not. Had the perceptual truths been *arbitrarily* different, our perceptual beliefs would not have been correspondingly different. In particular, had their contents been systematically false, because we are actually brains in vats, it seems that we still would have believed that we have hands. What we can arguably show, as Nozick [1981, Ch. 3] pointed out, is that our perceptual beliefs are *sensitive* – i.e., that had a typical one of them been false, we would not still have believed its content.

[18] This formulation assumes an ontologically-committal interpretation of set-theoretic claims. But, again, one can ask what beliefs we would have had had there been no non-vacuous set-theoretic truths, whether or not such truths turn on the existence of sets.

This suggests a way of refining *Answer 4*. Rather than requiring that we do not believe P in ~P worlds in which the set-theoretic truths are varied in any which way, we might just require that we do not believe P in the *closest* ~P worlds (in which we still determine whether P using the method that we actually used). In other words, rather than requiring that our set-theoretic beliefs are counterfactually persistent, we might require that we show that they are sensitive.

> *Answer 5 (Sensitivity)*: In order to "explain the reliability" of our set-theoretic beliefs it is necessary to show that, for any one of them, that P, that had it been the case that ~P, we would not still have believed P (using the method that we actually used to determine whether P).

Unfortunately, (b) remains false if (a) is true. Again, if our set-theoretic beliefs fail to be sensitive when interpreted as being about the "one true V", then they equally fail to be sensitive when interpreted as being about the pluriverse. More importantly, *Answer 5* still fails to satisfy (c). Even if, as a general rule, evidence that our belief that P is insensitive undermines that belief (which is highly controversial [White 2010, Sec. 4.1]), it is hard to maintain that evidence that our belief that P undermines that belief *when P is metaphysically necessary, if true*. The problem, again, is that virtually none of our beliefs in metaphysically necessary truths is *non-vacuously* sensitive. For instance, had the "bridge law" that atoms arranged chair-wise compose a chair been false, it seems that we still would have believed it. Moreover, if our beliefs in such bridge laws are undermined, then so too, it would seem, are our ordinary beliefs

that ascribe supervenient properties, such as that Jones is sitting in a chair (Clarke-Doane [2016a, 2.4]). This assumes a closure principle that one could conceivably deny. But it is hard to envision how our beliefs about chairs could be rationally insulated from our beliefs about the instantiation conditions of chairhood.

To sum up: *if* there is a reason to be a set theoretic pluralist, *then* it is not related to the challenge to establish a causal, explanatory, logical or even counterfactual dependence between our set-theoretic beliefs and the truths. Not only do none of *Answers 1 – 5* seem to satisfy both (a) and (b). None seem to satisfy (c). But there are two ways of having false beliefs of a kind, F. First, it could happen that the *F-truths* are different while our F-beliefs fail to be corresponding different. Second, it could happen that our *F-beliefs* are different while the F- truths fail to be correspondingly so. Even if the first possibility is inapt when the truths are set-theoretic, the latter remains. Perhaps there is an interpretation of "explain the reliability" related the latter such that (a) – (c) are all true.

5. Contingency

Despite Field's remarks on counterfactual persistence, he sometimes appears to conceive of Benacerraf's challenge as stemming from the possible variation of our beliefs. He writes,

> [Pluralists] solve the [Benacerraf] problem by articulating views on which though mathematical objects are mind independent, any view we had had of them would have been correct…[2005, 78].

Keeping in mind the need to relativize to methods of belief formation, let us interpret Field's proposal as follows.

> *Answer 6 (Failsafety)*: In order to "explain the reliability" of our set-theoretic beliefs it is necessary to establish, for any one of them, that P, that had we believed ~P, our belief still would have been true (had we used the method that we actually used to determine whether P).

As before, (a) is plausible. If our actual set-theoretic beliefs are true, then, had they been different, they would have been false – given that universalism is true. Actually, we must add that the set-theoretic truths are the same in the nearest worlds in which our beliefs are different. But this is immediate if the mathematical truths are metaphysically necessary, and it is plausible even if they are not. However, (b) is false. Even the pluralist concedes that had our set-theoretic beliefs been different because they were *inconsistent*, they would have been false. Had we believed ZFC + "there *is* a universal set", our set-theoretic beliefs would have been false – even assuming pluralism.[19]

This suggests the following weakening of Answer 6.

[19] I will mention a more fundamental problem with this proposal and the next shortly.

Answer 7 (Consistent Failsafety): In order to "explain the reliability" of our set-theoretic beliefs it is necessary to establish, for any one of them, that P, that had we believed ~P, but our set-theoretic beliefs remained consistent, then they our belief that ~P still would have been true (had we used the method that we actually used to determine whether P).

Components (a) and (b) are both now plausible. The problem is that (c) is false when understood in accord with either *Answer 6 or 7*. After all, had our perceptual beliefs been *very* different, but still consistent, a given one of them may have been false too. Had we had consistent perceptual beliefs as of goblins, for example, our belief that there is one still would have been false – since the closest worlds in which these conditions are met are worlds in which we are deluded somehow.

One could try to avoid this problem by requiring that our set-theoretic beliefs would have been true had they been different, consistent, and "relevantly similar". But absent an independent explication of relevant similarity, this proposal is without clear content.

Answer 6 and *Answer 7* falter because they equate a condition that is constitutive of realism, as traditionally conceived, with an epistemological problem. To say that had the truths been different, our beliefs would have been false is just to say that the truths do not counterfactually depend on our beliefs. In order to turn this condition into a problem, we must add that *we could have easily had different beliefs*. It then follows that we could have easily had *false* such beliefs.

Keeping in mind the need to relativize to methods of belief formation, this suggests a final interpretation of "explain the reliability".

> *Answer 8 (Safety)*: In order to "explain the reliability" of our set-theoretic beliefs it is necessary to establish, for any one of them, that P, that we could not have *easily* had a false belief as to whether P (using the method that we actually used to determine whether P).

Answer 8 is the only answer to the question with which we began of which I am aware that plausibly satisfies (c). Evidence that we could have easily had a false belief as to whether P (using the method that we actually used to form our belief) is a paradigm *undermining* (as opposed to rebutting) defeater of our belief that P. It gives us reason to give up our belief that P, but not by giving us "direct" reason to believe that ~P. Indeed, this seems to be how learning that our moral or religious beliefs are the products of evolutionary or social forces could be "debunking" (Mogensen [2016]). It could give us reason to think that we might have easily had different ones.

Moreover, (a) is plausible because the "set-theoretic orthodoxy" is apparently quite contingent. (One need not claim that the universalist's truths would be so contingent.) It is not hard to imagine scenarios in which different axioms were standardly accepted. As Hamkins writes,

> Imagine…that…the powerset size size axiom [(PSA) that for any x and y, |x| < |y| implies 2^x < 2^y] had been considered at the very beginning of set theory…and was subsequently added to the standard list of axioms. In this case, perhaps we would now look upon models of ~PSA as strange in some fundamental way, violating a basic intuitive principle of sets concerning the relative sizes of power sets; perhaps our reaction to these models would be like the current reaction some mathematicians (not all) have to models of ZF+~AC or to models of Aczel's anti-foundation axiom AFA, namely, the view that the models may be interesting mathematically and useful for a purpose, but ultimately they violate a basic principle of sets [2011, 19].

Nor can we merely imagine alternative set-theoretic axioms being accepted. There *actually are* many theorists who hold heretical views which do not seem to turn on any outstanding conjectures. For example, Boolos [1999, 121] does not accept all instances of Replacement, Potter [2004, Sec. IV] is skeptical of Choice, Rieger [2018, 17-18] rejects Foundation, and Friedman [FOM, 5.25.00] and Jensen [1995, 401] have sympathy for V = L. And while Koellner [2013, 21-22] suggests that there is "wide acceptance" of standard axioms, including "large" large cardinals, it certainly appears that history could have easily been otherwise. It is not as if alternatives to the orthodoxy foreclose standard applications of set theory to empirical science, for example. As Martin writes, "[f]or individual mathematicians, acceptance of an axiom is probably often the result of nothing more than knowing that it is a standard axiom [1998, 218]."[20]

[20] See also Maddy [1988]. In my [2016a, 2.3], I suggest that whatever contingency there is in our set-theoretic beliefs is due to the contingency of our abductive practices. I argue that our elementary mathematical beliefs may be modally robust, and that our set-theoretic beliefs may "best systematize" the former. If the contingency of our abductive practices is undermining, then all manner of our beliefs – mathematical and otherwise – are undermined. However, our "abductive practices" may not stand or fall together. What distinguishes those which lead us to accept

Whether (b) is true is uncertain. But we can at least see *how* it could be true. There is a *prima facie* case to be made that we were selected to have a reliable mechanism for deductive inference (Schechter [2010], [2013]).[21] Moreover, mathematics seems singularly devoted to rigorously applying that mechanism using languages invented to wear their "logical form" on their sleeves. So, if we have a reliable mechanism for deductive inference, mathematics would seem to afford the most favorable context for it to operate. Of course, we know from the undecidability of first-order logic that being able to reliably identify cases in which Q follows from P does not imply being able to reliably identify cases in which Q does *not* follow from P – including cases in which Q is a contradiction. Moreover, important set-theoretic principles – most famously, Naive Comprehension – have turned out to be inconsistent. But it is striking that such inconsistencies have generally been discovered promptly. I know of no set-theoretic principle which was once generally accepted by the set-theoretic community but later turned out to be inconsistent.

Strictly speaking, it may not be not enough for the pluralist to claim that we could not have easily had inconsistent set-theoretic beliefs. Again, unless the set-theoretic pluralist wishes to be a pluralist about consistency itself, she must accept a certain amount of arithmetic objectivity. However, just as it is arguable that we could not have easily had inconsistent set-theoretic beliefs, it arguable that we could not have easily believed that likes of ZF + ~Con(ZF). To believe such a theory would be to believe a theory, while also believing that it is inconsistent.

---

the Axiom of Choice from those which lead us to accept e = mc^2 may be precisely that the latter are more modally robust than the former. The latter, but not the former, are tested against a causally efficacious world.

[21] This is far from settled, however. For relevant discussion, see Cosmides and Tooby [1991].

Note that the pluralist *cannot* "explain the reliability" of our set-theoretic beliefs in the sense of *Answer 7* by establishing that, for any set-theoretic *proposition* that we believe, that P, had we believed ~P our belief still would have been true. Assuming that P holds in the closest world in which we believe that ~P, and that contradictions are impossible, *nobody* can establish that. (Indeed, this is another reason why neither *Answer 6* nor *Anwer 7* could satisfy (c).) Rather, the pluralist can arguably establish that, if P_{Choice} is the proposition expressed by the sentence, S_{Choice}, then, had we uttered ~S_{Choice}, we would have asserted a *~P_{Choice}-like proposition* -- where a *~P-like proposition* is the translation of ~P into a possibly distinct true proposition that intuitively shares ~P's "metaphysical content". If the Axiom of Choice, understood as a proposition, is true "in" some part of the pluriverse, then it is not false in another. But the pluralist holds that, along with sets, which are its subject matter, there are shmets. Shmets, we might say, are just like sets, except that some nonempty ones lack "SChoice functions".

To sum up: the claim that some beliefs are inconsistent (or Pi_1 unsound) *or* fail to be true in the "one true V" is weaker than the claim that some beliefs are inconsistent (Pi_1 unsound) simpliciter. In this sense, the pluralist is indeed better positioned than the universalist to argue that our set-theoretic beliefs are safe. However, whether (b) is true depends on the broadly empirical conjecture that our deductive practices are significantly less contingent than our (non-logical) set-theoretic beliefs. If this conjecture proves to be false, then, absent an alternative analysis of Benacerraf's challenge that satisfies (a) – (c), it is also false that the

pluralist is better positioned than the universalist to answer Benacerraf's challenge – contrary to what is widely assumed.

6. Conclusion

What could Benacerraf's challenge be such that the set-theoretic pluralist can answer that challenge, but the universalist cannot? I have argued that it could not be any of the challenges with which it has traditionally be identified by its advocates, such as Benacerraf and Field. In particular, the set-theoretic pluralist is no better positioned to answer the challenge to establish a causal, explanatory or counterfactual dependence between our beliefs and the truths. Moreover, none of these challenges seems to be undermining, if impossible to meet. However, I have argued that Benacerraf's challenge could be the challenge to show that our set-theoretic beliefs are *safe* – i.e., the challenge to show that we could not have easily had false ones (using the method that we actually used to form ours). It does appear impossible for the universalist to show that our set-theoretic beliefs are safe, and the apparent impossibility of showing this is arguably undermining. But whether the pluralist can show that our set-theoretic beliefs are safe is unclear. It depends on whether our set-theoretic beliefs are significantly more contingent than our deductive practices. If they are not, then, absent an alternative answer to the question that began this paper, it is unclear what the epistemological argument for pluralism could be.[22]

---

[22] It might be thought that I have overlooked a way of understanding the reliability challenge, a formulation occasionally pressed by Field himself. In his [1989], he writes,

> If the intelligibility of talk of "varying the facts" is challenged... it can easily be dropped without much loss to the problem: there is still the problem of explaining the *actual* correlation between our believing "p" and its being the case that p [238, italics in original].

I do not know what this means. It might be taken to involve showing that the correlation holds in nearby worlds, so the actual correlation is no "fluke". But, in that case, we are just back to something like safety or sensitivity. Perhaps, then, there is a hyperintensional sense of "explanation" according to which one can intelligibly request an

Bibliography

Balaguer, Mark. [1998] *Platonism and Anti-Platonism in Mathematics*. New York: Oxford University Press.

Barton, Neil. [2016] "Multiversism and Concepts of Set: How Much Relativism is Acceptable?" in Francesca Boccuni & Andrea Sereni (eds.), *Objectivity, Realism, and Proof*. Springer. 189-209.

Beall, J.C. [1999] "From Full-Blooded Platonism to Really Full-Blooded Platonism." *Philosophia Mathematica*. Vol. 7. 322—327.

Benacerraf, Paul. [1973] "Mathematical Truth." *Journal of Philosophy*. Vol. 60. 661 – 679.

Chihara, Charles. [1990] *Constructability and Mathematical Existence*. Oxford: Oxford University Press.

Clarke-Doane, Justin. [2012] "Morality and Mathematics: The Evolutionary Challenge." *Ethics*. Vol. 122. 313-340.

explanation of the "merely actual correlation" between our beliefs and the truths. But if that sense is not given by *Answer 2* or *3*, then I am not sure what it is. Even if there were such a sense, it is unclear how the apparent impossibility of offering such an explanation could *undermine* our set-theoretic beliefs, for reasons surveyed in Section 3. (Even if we cannot explain the "merely actual correlation" between our moral or mathematical beliefs and the truths, in some hyperintensional sense of that phrase, we might still be able to show that those beliefs are sensitive, safe (and objectively probable), realistically construed.) Finally, even if the above challenge can be made out, is distinct from those surveyed, and is worth taking seriously, whether the pluralist is better positioned to answer it would itself seem to depend on whether our non-logical set-theoretic beliefs are significantly more contingent than our deductive practices.

-----. [2016] “What is the Benacerraf Problem?” in Fabrice Pataut (ed.), *New Perspectives on the Philosophy of Paul Benacerraf: Truth, Objects, Infinity.* Dordrecht: Springer.

-----. [Forthcoming] *Morality and Mathematics.* Oxford: Oxford University Press.

Colyvan, M. 2007. “Mathematical Recreation versus Mathematical Knowledge,” in M. Leng, A. Paseau, and M. Potter (eds.), *Mathematical Knowledge*. Oxford: Oxford University Press, 109–22.

Cosmides, Leda and John Tooby. [1991] “Reasoning and Natural Selection.” *Encyclopedia of Human Biology, vol. 6.* San Diego: Academic Press.

Enoch, David. [2010] “The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It.” *Philosophical Studies.* Vol. 148. 413-438.

Field, Hartry. [1989] *Realism, Mathematics, and Modality*. Oxford: Blackwell.

-----. [2005]. “Recent Debates about the A Priori,” in T. S. Gendler and J. Hawthorne (eds.), *Oxford Studies in Epistemology, Volume 1*. Oxford: Oxford University Press, 69–88.

-----. [1998] “Which Mathematical Undecidables Have Determinate Truth-Values?” in Dales, H. Garth and Gianluigi Oliveri (ed.), *Truth in Mathematics*. Oxford: Oxford University Press. 291–310

Gödel, Kurt. [1947] “What is Cantor’s Continuum Problem?” Revised and reprinted in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics*. Englewood Cliffs, NJ: Prentice-Hall, 1964.

Goldman, Alvin. [1967] “A Causal Theory of Knowing.” *Journal of Philosophy*. Vol. 64. 357—372.

Hamkins, Joel David. [2011] "The Set-Theoretic Multiverse". arXiv. Available at: https://arxiv.org/abs/1108.4223

-----. [2012] "The Set-theoretic Multiverse." (revised version) *Review of Symbolic Logic*. Vol. 5. 416-449.

Hellman, Geoffrey. [1989] *Mathematics Without Numbers*. Oxford: Oxford University Press.

Jensen, Ronald. [1995] "Inner Models and Large Cardinals." *Bulletin of Symbolic Logic*. Vol. 1. 393 -- 407.

Joyce, Richard. [2008] "Precis of Evolution of Morality and Reply to Critics." *Philosophy and Phenomenogical Research*. Vol. 77. 213 – 67

Koellner, Peter. [2013] "Hamkins on the Pluriverse." Manuscript.

Leng, Mary. [2009] ""Algebraic" Approaches to Mathematics." in Otavio Bueno and Oystein Linnebo, *New Waves in the Philosophy of Mathematics.* New York: Palgrave Macmillan.

-----. [2010] *Mathematics and Reality.* Oxford: Oxford University Press.

Liggins, David. [2006] "Is there a Good Epistemological Argument against Platonism?" *Analysis*. Vol. 66. 135–141.

Linnebo, Øystein. [2006] "Epistemological Challenges to Mathematical Platonism." *Philosophical Studies*. Vol. 129. 545–574.

Linsky, Bernard and Edward Zalta. [1995] "Naturalized Platonism versus Platonism Naturalized." *Journal of Philosophy*. Vol. 92. 525—555.

Maddy, Penelope. [1988] "Believing the Axioms I & II." *Journal of Symbolic Logic.* Vol. 53. 481—511 & 763—764.

Martin, Tony. [1976] “Hilbert's First Problem: The Continuum Hypothesis.” in Browder, Felix (ed.), *Mathematical Developments Arising from Hilbert Problems (Proceedings of Symposia in Pure Mathematics. Vol. 28)*. Providence, Rhode Island: American Mathematical Society.

McFedridge, I.G. [1990] “Logical Necessity.” in John Holdane and Roger Scruton (eds.), *Logical Necessity and Other Essays*. London: Aristotelian Society.

Mogensen, Andreas. [2016] “Contingency Anxiety and the Epistemology of Disagreement.” *Pacific Philosophical Quarterly*. Vol. 97. 590–611

Nozick, Robert. [1981] *Philosophical Explanations*. Oxford: Oxford University Press.

Potter, Michael. [2004] *Set Theory and Its Philosophy*. Cambridge: Cambridge University Press.

Pritchard, Duncan. [2008] “Safety-Based Epistemology: Whither Now?,” *Journal of Philosophical Research*. Vol. 34. 33–45.

Putnam, Hilary. [1980] “Models and Reality.” *Journal of Symbolic Logic*. Vol. 45. 464—482.

Rieger, Adam. [2011] “Paradox, ZF, and the Axiom of Foundation.” in DeVidi, David, Hallet, Michael, and Peter Clark (Eds.), *Logic, Mathematics, Philosophy, Vintage Enthusiasms: Essays in Honour of John L. Bell* (The Western Ontario Series in Philosophy of Science). New York: Springer.

Sayre-Mccord, Geoffrey. [1988] “Moral Theory and Explanatory Impotence.” *Midwest Studies in Philosophy*. Vol. XII. 433 – 457.

Schechter, Joshua. [2010] “The Reliability Challenge and the Epistemology of Logic.” *Philosophical Perspectives*. Vol. 24. 437-464.

-----. [2013] “Could Evolution Explain our Reliability about Logic?” in Tamar Szabo Gendler & John Hawthorne (eds.), *Oxford Studies in Epistemology, Vol. 4*.

Shapiro, Stewart. [2000] *Philosophy of Mathematics: Structure and Ontology*. New York: Oxford University Press.

Steiner, Michael. [1973] "Platonism and the Causal Theory of Knowledge." *Journal of Philosophy*. Vol. 70. 57–66.

Street, Sharon. [2008] "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About." *Philosophical Issues*. Vol. 18. 207 – 228

Warren, Jared. [2015] "Conventionalism, Consistency, and Consistency Sentences." *Synthese*. Vol. 192. 1351-1371.

White, Roger. [2010] "You Just Believe that Because." *Philosophical Perspectives.* Vol. 24. 573 – 612.