

UNPICKING PRIEST'S BOOTSTRAPS

(Philosophy of Mathematics)

ABSTRACT. Graham Priest has argued that the fruits of classical set theory can be obtained by naive means through a puzzling piece of reasoning often known as the *bootstrapping argument* [Priest, 2006]. I will demonstrate that the bootstrapping involved is best understood as viciously circular and thus, that these fruits remain forbidden. The argument has only one rehearsal in print and it is quite subtle. This paper provides reconstruction of the argument based on [Priest, 2006] and attempts some fixes and alternative construals to get around some elementary problems. Despite these efforts the argument remains unconvincing.

1. INTRODUCTION

The bootstrapping argument makes use of a pleasing model-theoretic construction the highlights of which we briefly sketch. Let *naive set theory*, abbreviated *NST*, be the conjunction of the axioms of extensionality and naive comprehension articulated in the underlying paraconsistent logic *LP* [Priest, 1979].¹ The use of just these axioms is seen as attractive as both of them have a strong intuitive pull and are arguably constitutive of the concept of set. Extensionality tells us that any two sets with the same members are, in fact, one and the same. Naive comprehension tells us that every definition of a collection gives us a set. The latter principle is inconsistent in classical logic, hence the use of a paraconsistent logic *LP*. An *LP*-model is simply a model of the non-classical logic *LP* [Priest, 2008]. Here is the result:

Proposition 1. (Priest) *Suppose that \mathcal{M} is a classical model of ZF containing two inaccessible cardinals κ_1 and κ_2 . Then there is an LP -model $\mathcal{M}^\sim = \langle D^\sim, I^\sim \rangle$ such that:*

- (1) \mathcal{M}^\sim is a model of $ZF + NST$; and
- (2) \mathcal{M}^\sim contains a model \mathcal{N} where \mathcal{N} is a classical model of ZF .²

Informally speaking, this tells us the given an ordinary model of ZF , one may obtain a non-classical model of $ZF + NST$ which *contains* a classical model of ZF : it is an interesting result. The basic strategy for the proof

I would like to thank Zach Weber for his helpful comments on an early draft of this paper.

¹It should be noted that Priest (and others) have considered logical frameworks other than *LP* for the purposes of a naive set theory. For example, non-monotonic *LP* and logics with weak relevant conditionals. Such logics are usually able to prove more than *LP*, which is very weak, and are thus more useful for mathematics. That said, the bootstrapping argument is specifically concerned with a naive set theory based on *LP* and as such, this paper predominantly focus upon it.

²I have stated the result in much the same way as Priest, however, it is not optimal. Using standard techniques, it can be seen that one inaccessible suffices for Priest's result (see Lemma 8.2 in [Kunen [2006], p141]).

is to take the model \mathcal{M} and use that to define a model \mathcal{M}^{\sim} which has the requisite properties. We don't need to concern ourselves with any more detail than this other than to note that the result is established using the ordinary tools of classical ZF . This will be important later.

Priest then argues that this puts the naive set theorist in a position to recapture or make use of classical set theory. This would be a desirable position for the naive set theorist. They could claim to have a non-classical foundation for mathematics but still be able to use the well understood tools of classical ZF . He makes the following remarks:

... to the question of what to make of the theorems of orthodox set theory, ZF , on this approach. The answer is obvious. Since the universe of sets is a model of ZF (as well as naive set theory), these hold in it. We may therefore establish things in ZF in the standard classical way, knowing that they are perfectly acceptable from a paraconsistent perspective. We cannot, of course, require the theorems of ZF to be consistently true in that universe; but if, on an occasion, we do require a consistent interpretation of ZF , we know how to obtain this too. The universe of sets has a consistent substructure that is a model of ZF . [Priest, 2006, p257-258]

The big question here is: what should the naive set theorist make of the orthodox set theory ZF ?³ He claims that we can have our cake and eat it on the basis of the following argument:

Claim 2. The universe of sets is a model of ZF as well as naive set theory.

Claim 3. Hence, the theorems of ZF are perfectly acceptable from a paraconsistent perspective.

Claim 4. Thus, the naive set theorist can establish claims in ZF in the standard classical way.

I believe Claim 2 is intended to be understood as following from Proposition 1. As I understand it, Claim 3 is supposed to follow from Claim 2 since Claim 2 tells us that the universe of sets is a model of ZF . Moreover, if we want to go further and demand consistency, we could work within the classical model \mathcal{N} of ZF given by Proposition 1.

The rest of the paper is divided as follows. In the next section, we examine the argument for Claim 2 and find its justification wanting. In Section 3, we examine the prospects of establishing Claim 3 in the absence of Claim 2. Finally in Section 4, we reexamine the argument for Claim 2 in the light of the previous section and find that even we turn a blind eye to its weak justification, the argument from Claim 2 to Claim 3 rests on circular reasoning.

³In deference to Priest, we'll stay in ZF although standard foundation practice includes the axiom of choice and thus, ZFC .

2. THE UNIVERSE OF SETS IS A MODEL OF ZF AS WELL AS NAIVE SET THEORY

As has been noted before, the argument from Proposition 1 to Claim 2 is troubling: at best, it is underdone [Weir, 2004, Young, 2014]. The proof of Proposition 1 shows how to construct a particular model \mathcal{M}^\sim with the desired properties given a classical model \mathcal{M} containing inaccessible cardinals κ_1 and κ_2 . On this basis, Priest claims:

We may therefore suppose that the true interpretation of the language of set theory has these properties. [Priest, 2006, p257]

From here it would be easy to obtain Claim 2, however at first blush, this seems far too bold. Just because we have a particular interpretation of NST which contains a model of ZF , this does not tell us that the true interpretation of set theory enjoys this property. To emphasise this point, we should note that if there are any models of ZF , then there are ill-founded models of ZF .⁴ The received interpretation of classical ZF tells us that the true interpretation of the membership relation is well-founded. Thus, the existence of this *pathological*, ill-founded model does not tell us that membership is really ill-founded. Similarly the existence of \mathcal{M}^\sim doesn't give us much reason to think that the true interpretation of membership is like that of \mathcal{M}^\sim . All we've learned is that this possibility has not been ruled out. Priest is aware of this problem and states:

"Specifying the correct interpretation is always a further issue. The model \mathcal{M}^\sim at least suffices to demonstrate the possibility of interpretations of naive set theory that have the above properties." [Priest, 2006, p257]

But this still leaves open the question: what positive reason do we have to think that the true interpretation of set theory is like \mathcal{M}^\sim ? This is all we get:

"This is an appealing picture. The cumulative hierarchy (up to κ_1) is a perfectly good, consistent set theoretic structure; but it does not exhaust the universe of sets. There may be non-well-founded sets (such as the set of all sets) and inconsistent sets, such as the set of all sets that are not members of themselves." [Priest, 2006, p257]

So the best reason we are given for accepting that the universe of sets is like \mathcal{M}^\sim is the *congeniality* of that assumption. Given the load this claim bears in the overall argument, I think it would be fair to say that this defence is insufficient and stop here. However, we shall see later that even if we grant Priest leave to provide a better argument, a deeper trap lies ahead.

⁴A simple argument using the compactness theorem suffices for this purpose.

3. THE THEOREMS OF ZF ARE PERFECTLY ACCEPTABLE FROM A PARACONSISTENT PERSPECTIVE

Given the poor defence provided for Claim 2, we might wonder whether Claim 3 can still be defended in its absence. Can we argue that ZF is acceptable from a paraconsistent perspective using Proposition 1, but without going through Claim 2? If we look more closely at Proposition 1, we see that it is essentially a relative consistency proof. Among other things, it tells us that if ZF is consistent (i.e., has a model), then the theory $ZF + NST$ is non-trivial (i.e., has a model in which not everything is true). At best, this result tells us from the classical (not paraconsistent) perspective, that NST is sensible, in the sense of being non-trivial. Loosely speaking, it tells us that if ZF is coherent (in a very weak sense), then so is $ZF + NST$. This is a nice result. However, this is not Priest's goal: for his purposes, he wants something like the converse. He wants to know that ZF is consistent from the paraconsistent point of view. Thus, he wants something like a solution to the following problem:

Problem 5. If there is a non-trivial LP -model of NST , then there is a classical model of ZF .⁵

Informally speaking, a positive answer to this problem would tell us that if NST is coherent (in the weak sense of being non-trivial), then so is classical ZF . This would tell us that from the paraconsistent perspective, classical ZF makes a minimal amount of sense. It doesn't tell us exactly what to make of the theorems of ZF , but it's a good start. Ideally, the naive set theorist would solve this problem using only the tools afforded by NST .⁶

One might be tempted to think a solution to Problem 5 is offered by the model \mathcal{M}^\sim . \mathcal{M}^\sim is a model of NST which contains a model \mathcal{N} of classical ZF . It could appear that we simply obtained the classical model \mathcal{N} from the LP -model \mathcal{M}^\sim . This, however, is fool's gold. Proposition 1 begins with the assumption that we already have a model of $ZF + 2$ inaccessible cardinals from which we gain a model of NST which contains a model of ZF . So we don't really get our classical model of ZF from the LP -model but rather the original model of $ZF + 2$ inaccessible cardinals. More formally, we might put this as follows:

Fact 6. $Con(ZF + 2 \text{ inaccessible}) \rightarrow Con(ZF)$.⁷

We start by assuming there is a model of ZF which contains two inaccessible cardinals and we then obtain a model of ZF . In this shape, the result is an obvious triviality. But worse still, Problem 5 has a negative answer

⁵By a non-trivial LP -model, we mean a model in which at least one sentence is not both true and false.

⁶If NST is to live up to its foundational role, then it is important that it is strong enough to establish such claims and that we do not resort to classical ZFC (at least until we've established that such a move is acceptable).

⁷We follow standard set theoretic conventions here in the use of predicate, Con . $Con(ZF)$ can be understood as saying that there is no proof of $0 = 1$ using the axioms of ZF . We shall take it that this has been formalised in the language of set theory. For a more precise account of how to do this see Boolos's [1993, p44]. However, we shall often exploit the completeness theorem and take $Con(\Gamma)$ to say that Γ is satisfiable. This is always acceptable if our latent metatheory is ZF .

if we address it in classical ZF . The argument for 1 on page 256 can easily be seen to establish the following general result:

Fact 7. *If \mathcal{M} is a well-founded classical model of extensionality where for all $x, y \in \mathcal{M}$ there is some $z \in \mathcal{M}$ such that $x, y \in z$, then there is a non-trivial LP-model \mathcal{M}^\sim of extensionality and naive comprehension: i.e., NST .*

It is easy to see that the proof of this fact can be carried out in classical ZF (and indeed much weaker theories). Moreover, classical ZF can easily prove that there is a well-founded model of extensionality satisfying the condition above, which means that it can also show that there is a non-trivial model of NST .⁸ But then if Problem 5 had a positive solution, we'd be able to prove $Con(ZF)$ using ZF alone; which contradicts Gödel's second incompleteness theorem. We thus see that in the absence of Claim 2, it is not clear how we can get to Claim 3: it is not at all clear that ZF is acceptable from the paraconsistent perspective. This highlights the important role Claim 2 plays in Priest's argument.

At this point we could consider an alternative goal for Priest. At heart, the bootstrapping argument was intended to address the vague question of what to make of classical ZF from the paraconsistent perspective. We just argued that without Claim 2 there is little reason to think that ZF is acceptable from the paraconsistent perspective. But perhaps NST could *recapture* the theorems of ZF . While the naive set theorist may still doubt their reliability, at least they could replicate the workings of the classical set theorist. A positive solution to the following problem would go some way to realising this goal:

Problem 8. Is there some translation τ of the language of set theory into itself (for example we might try binding quantifiers to a particular model) such that:

- (1) if $ZF \vdash A$, then $NST \vdash \tau(A)$; and
- (2) for some B , $NST \not\vdash \tau(B)$.

The first condition sets up the recapturing of classical mathematics. We add the second restriction so that we are forced to avoid trivial translations. Otherwise, for example, we could relativise every sentence to the set of objects, x , which are such that every set is: identical and not identical to x ; a member of and not a member of x ; and has x as a member and doesn't have x as a member. Using naive comprehension it can be seen that such a set exists in any model of NST , from which one can show by induction that $NST \vdash \tau(B)$ for all sentences B .⁹ That would be a very poor kind of recapture.

⁸Simply, let \mathcal{M} be $\langle \omega, \in \rangle$ and define \mathcal{M}^\sim by letting a be, say, 5. See [Priest 2006, p256].

⁹We make use of the completeness theorem for LP here.

However, a solution to this problem would also clash with Gödel's second incompleteness theorem. To see this, suppose that a positive solution to this problem had been found using the lights of classical ZF . Now fix some B such that $NST \not\vdash \tau(B)$. Then by (1), we see that $ZF \not\vdash B$. By the completeness theorem for first order logic, we may then obtain a model \mathcal{M} of $ZF \cup \{B\}$. But this would again mean that we could establish the consistency of ZF using ZF .

3.1. A remark on metatheory. In both of these Gödelian responses, I have used an *external*, classical metatheory. Given that I think it would be ideal to address these problems from the *internal* perspective of the NST framework, this warrants some further remarks. I think consideration of the internal approach raises some new and interesting challenges the solutions to which would also be a major step forward for non-classical set theory. A thorough investigation of this avenue is outside the scope of this paper, however, this presents no problem for us in that we are only trying to show that Priest's argument is incomplete.

Nonetheless, let us consider what it would be like to address Problem 5 in NST . There is an immediate issue in that it isn't clear that a classical model as defined in NST is particularly classical, hence it is not clear that a solution to the problem would demonstrate what it purports to show at first blush. While it might be possible to define a bivalent model-theoretic apparatus within the NST framework, it's not at all clear that this is *really* bivalent.¹⁰

Given the well-known weakness of LP , we may also have some cause to doubt whether such a result could be established in NST .¹¹ It would be ideal to show that there can be no such proof, but in the absence of a well-worked out way of doing metatheory in NST , it is difficult to know how to establish such a result.¹² But suppose, for the sake of argument, that an internal solution to Problem 5 could be provided. Then the Gödelian arguments above draws out a significant tension between NST and classical ZF . While our goal was to be able to take up the classical perspective of set theory in an fashion that is acceptable for the paraconsistent set

¹⁰For example, suppose we have an LP model \mathcal{M} of NST which contains some \mathcal{N} which \mathcal{M} says is a classical model. Then all this tells us is that

$$v_{\mathcal{N}}(\forall \varphi (\mathcal{M} \models \varphi \Leftrightarrow \mathcal{M} \not\models \neg \varphi)) \in \{\{1\}, \{0, 1\}\}.$$

However, it could still be the case that there is some φ such that $v_{\mathcal{N}}(\mathcal{M} \models \varphi) = 1$ and yet $v_{\mathcal{N}}(\mathcal{M} \models \neg \varphi) = 1$.

¹¹This might give us reason to abandon LP as our foundational logic and opt for something stronger, perhaps a relevant conditional. Such a path takes us further afield from Priest's original motivations for the bootstrapping argument. If we do take this path, then a revision to Problem 5 is warranted. Since we have revised to a new foundational logic, \mathcal{L} , we are now interested in non-trivial models of naive comprehension and extensionality over the new logic: call this $NST_{\mathcal{L}}$. This then raises an interesting, new problem:

Problem 9. Is there a logic \mathcal{L} such that $NST_{\mathcal{L}}$ is non-trivial and such that $NST_{\mathcal{L}}$ can prove that:

If $NST_{\mathcal{L}}$ has a non-trivial model, then ZF has a classical model.

This looks like a difficult problem. A positive answer could well provide us with some good reasons to accept paraconsistent set theory. For the purposes of this paper I am content to leave it open.

¹²Presumably, we'd want to construct a counter-model, but this would appear to require technical tools which are available in classical ZF but to which we lack access in NST .

theorist, there is something awry in this situation from the classical perspective. On the one hand we know that classical ZF is stronger than NST in that it is able to provide a non-trivial model for it. On the other hand, ZF seems weaker than NST given that there are simple model-theoretic propositions that are provable in NST but not in ZF . In the absence of a clearly articulated NST version of Problem 5 it is not possible to properly sharpen this point, but it would seem odd to claim to vindicate the classical perspective while at the same time proving something that the classical perspective says is out of bounds.

4. A DEEPER PROBLEM

Recall that the goal of naive set theory, as construed here, is to be able to provide a foundation for mathematics with the simple axioms of NST and the paraconsistent logic LP . The goal of the bootstrapping argument is to show that one can take up naive set theory but still use the better understood and more user-friendly ZF . We have seen that the argument for this contention rests on Claim 2: that the true interpretation of the language of set theory satisfies $ZF + NST$. The argument for Claim 2 rests on two further premises:

Claim 10. Proposition 1 gives us a model \mathcal{M}^\sim of NST which contains a classical model of ZF .

Claim 11. \mathcal{M}^\sim provides a compelling illustration of what the universe of sets ought to be like.

As we have seen already, the arguments for Claim 11 are thin at best. However, for the sake of argument, let us suppose that they can be filled out. A further problem remains. First we note that Claim 10 is playing a crucial role. We are only brought into position to countenance the intelligibility of Claim 2 on the basis of the model \mathcal{M}^\sim given to us by Claim 10. Proposition 1 tells us that the possibility of Claim 2 holding has not been ruled out. Without Claim 10, Claim 2 is little more than wishful thinking. Moreover, the positive argument for Claim 2 given by Claim 11 is only of value once the intelligibility of Claim 2 has been established using Claim 10. This is why Priest takes such pains to demonstrate Proposition 1.

But now recall that Proposition 1 was established using the tools of classical ZF . In ordinary practice, there would be nothing wrong with this move. However, here we have a problem. Claims 10 and 11 are premises in an argument for Claim 2, which is then used to establish Claim 3: that the theorems of classical ZF are acceptable from a paraconsistent perspective. But in order to establish the proposition behind Claim 10, we used classical ZF . We thus assumed from the outset that classical ZF was acceptable from the paraconsistent perspective. The conclusion that Priest wants to draw from this argument was assumed to be true at the beginning. This is a textbook case of question begging: Priest's conclusion is one of his premises!¹³

¹³It is worth noting that we have only considered the possibility of establishing Proposition 1 using the classical ZF techniques adopted by Priest. Moreover, it seems clear that Priest believes this external strategy is sufficient: "... the argument constructing \mathcal{M}^\sim above can

To give it a slightly more informal gloss, we were trying to establish the credentials of classical ZF from a paraconsistent perspective. We wanted show that even if we adhere to a non-classical foundation of naive set theory, we can still make use of the tools of classical ZF . This was our goal. But only after such an argument has been successfully executed, could we be licensed to use classical ZF . Priest, however, has broken the rules and is already making use of classical ZF to establish a crucial premise in the argument toward this goal. Unless Proposition 1 can be established using NST , and we been given no reason to think that it can be, we are forced to see this as a clear piece of circular reasoning.

REFERENCES

- George Boolos. *The Logic of Provability*. Cambridge University Press, 1993.
- Kenneth Kunen. *Set Theory: an introduction to independence proofs*. Elsevier, Sydney, 2006.
- Graham Priest. The logic of paradox. *Journal of Philosophical Logic*, 8:219–241, 1979.
- Graham Priest. *In Contradiction: A Study of the Transconsistent*, volume 25. Oxford University Press, 2006.
- Graham Priest. *An Introduction to Non-Classical Logic: From If to Is*. Cambridge University Press, Melbourne, 2008.
- Alan Weir. There are no true contradictions. In Graham Priest, J. C. Beall, and Bradley Armour-Garb, editors, *The Law of Non-Contradiction*. Clarendon Press, 2004.
- Gareth Young. *Revenge: Dialetheism and its Expressive Limitations*. PhD thesis, University of Glasgow, 2014.

be carried out in ZF , and so is perfectly acceptable” [Priest 2006, p258]. But perhaps we can prove it *internally* using NST alone (or some other non-classical logic). For similar reasons to those discussed in Section 3.1, I am reluctant to raise any arguments against such a possibility and indeed it seems like an interesting question. That said, we may have some reason to be pessimistic about the value of an *internal* solution. Proposition 1 makes specific mention of providing a model of NST which contains a classical model. From the external classical perspective, we have good reason to think that the model provided is genuinely classical. However, as we discussed in Section 3.1, the kinds of model that NST thinks are classical could well be non-classical as well. As such, it seems unlikely that the internal version of Proposition 1 could help us in establishing Claim 10 and from that Claim 2.