# RESPONSE TO MY CRITICS

Roy T. Cook
University of Minnesota

## Abstract

During the Winter of 2011 I visited SADAF and gave a series of talks based on the central chapters of my manuscript on the Yablo paradox. The following year, I visited again, and was pleased and honored to find out that Eduardo Barrio and six of his students had written 'responses' that addressed the claims and arguments found in the manuscript, as well as explored new directions in which to take the ideas and themes found there. These comments reflect my thoughts on these responses (also collected in this issue), as well as my thoughts on further issues that arose during the symposium that was based on the papers and during the many hours I spent talking and working with Eduardo and his students.

KEY WORDS: The Yablo Paradox; Circularity; Anti-Realism: Non-well-founded Sets; Paradoxes.

## Resumen

Durante el invierno de 2011, visité SADAF y dí una serie de conferencias sobre los capítulos centrales de mi manuscrito sobre la paradoja de Yablo. El año siguiente, visité Buenos Aires nuevamente y tuve el placer y el honor de descubrir que Eduardo Barrio y seis de sus estudiantes habían escrito respuestas que abordan las afirmaciones y argumentos de mi manuscrito, además de explorar nuevas direcciones en las cuales considerar las ideas y temas encontrados allí. El presente trabajo refleja mis pensamientos sobre estas respuestas (también incluidas en este volumen), así como mis ideas sobre otras cuestiones que surgieron durante el simposio en el que se presentaron los artículos y en las muchas horas que nosotros discutimos y trabajamos con Eduardo y sus estudiantes.

PALABRAS CLAVE: La Paradoja de Yablo; Circularidad; Antirrealismo; Conjuntos infundados; Paradojas.

## 1. Introduction

I shall not have much to say regarding the content of Professor Barrio's contribution, since it in a very straightforward sense speaks for itself, and my silence regarding the details of its content says enough – Barrio's contribution is meant to be, for the most part, a detailed overview of the state of play with regard to the Yablo paradox (and, in particular, my own work on this puzzle), and it carries out this task admirably. The fact that the contributions that follow, and my responses to them, can dive immediately into the substantial issues is due in no small part to Barrio's admirable job in laying out the philosophical landscape in his introductory contribution.

I will note, however, that I find Barrio's final observation –that the arithmetic version of the Yablo paradox formulated as an ω-sequence of individual sentences is ω-inconsistent– and his further observation –that ω-inconsistency is bad enough to justify the term "paradox", regardless of proof-theoretic consistency of the sentences in question– to be on the mark. In fact, I firmly believe that, in some substantial sense of "circular", the original Yablo paradox is not circular, but it is difficult to argue conclusively for this claim with the inadequate accounts of circularity that are currently on offer. Hence the infinitary construction, which can be proven to be non-circular in a precise sense. Thus, in broad outline at least, I agree with Barrio that the original version is (probably) non-circular, and that it is a genuine paradox (and I suspect that we agree about a great deal more!)

This is all I will say about Barrio's contribution. But it would be inappropriate to move on and address the other papers collected here without first thanking Barrio for other efforts that will not be evident merely from reading his excellent introduction to this symposium. The very existence of this symposium, and of the excellent short essays to which I respond below, owes much to Barrio's hard work (and the equally impressive efforts of his students). The dialogue between myself and the six students whose essays appear here began when Barrio invited me to visit SADAF in the summer of 2011. During this period I presented a draft of my forthcoming book on the Yablo paradox (*The Yablo Paradox: An Essay on Circularity*, Oxford), and was both pleased and impressed with the level of careful and creative feedback on this material that these students (and others) so generously offered during that time. The final version of the book is much better than it would otherwise have been, and I would like to take this opportunity to thank all those who struggled with me on this topic during that visit. But the debts owed do not stop there, since a year later Barrio invited me back to Buenos Aires for a second visit, in order to respond to the papers collected here. Again, the generous and collegial atmosphere evident to me throughout this visit is, I hope, also evident to the reader in the discussion collected here. I have once again thoroughly enjoyed my time spent in Argentina, and I hope that opportunities for further visits –and thus further opportunities to thank Barrio and his students– arise in the future.

## 2. Response to Picollo

In Picollo's essay "The Old-Fashioned Yablo Paradox", the main task is to formulate a criterion for circularity that (i) intuitively captures (at least most of) our intuitions regarding what sentences are (and are not) circular, and (ii) judges (at least one version of) the first-order arithmetical version of the Yablo paradox to be non-circular. Before leveling criticisms, it is worth noting that I agree with Picollo that such a criterion is lacking, and is sorely needed –one of the central tasks of the book was to survey, and ultimately reject, various attempts to give such an account of circularity. In addition, I think that Picollo's general strategy– which focuses on the syntax of the statements involved rather than on less-fine-grained semantic or inference-based notions –is in many ways on the right track. I don't think, however, that the account she provides here does the job. There are two problems with her approach– one is simple and easy to fix, but the second is more substantial.

Picollo's proposed definition of mention-circularity (or m-circularity) is comprised of three successive definitions:[1]

> **Definition 1**: $\Phi$ directly m-refers to $\Psi$ if and only if $\Phi$ contains a singular term $t$ and $t = \langle\Phi\rangle$.
> **Definition 2**: $\Phi$ m-refers to $\Psi$ if and only if the ordered pair $\langle\Phi, \Psi\rangle$ belongs to the transitive closure of direct m-reference.
> **Definition 3**: $\Phi$ is m-circular if and only if it contains a singular term $t$ such that $t = \langle\Psi\rangle$ and $\Psi$ m-refers to itself.

The first problem with this set of definitions is that it seems to judge certain simple constructions incorrectly. Let $\Phi_1$, $\Phi_2$, and $\Phi_3$ be three distinct primitive terms, $\Psi_1$, $\Psi_2$, and $\Psi_3$ three distinct primitive predicates, and stipulate that:

$$\Phi_1 = \text{“}\Psi_1(\Phi_1)\text{”}$$
$$\Phi_2 = \text{“}\Psi_2(\Phi_1)\text{”}$$
$$\Phi_3 = \text{“}\Psi_3(\Phi_2)\text{”}$$

Obviously, $\Phi_1$ is m-circular, and less obviously $\Phi_2$ is m-circular on Picollo's definition: $\Phi_2$ contains the term $\Phi_1$ which m-refers to itself. $\Phi_3$,

[1] I ignore the additional definition specifying when a set of sentences is circular, since, although important in general, it is irrelevant to my comments here.

however, contrary to intuition, is not m-circular, since $\Phi_2$ does not m-refer to itself.

This apparent counterexample is easily avoided, however, by a simple fix. The intuitive idea behind the thought that $\Phi_3$ is circular is that it refers to a sentence that refers to a sentence that is circular (even if it does not directly refer to an m-circular sentence). Hence, we can obtain the desired result by a simple reformulation of Picollo's Definition 3:

> **Definition 3.5**: $\Phi$ is m-circular if and only if it m-refers to some formula $\Psi$ such that $\Psi$ m-refers to itself.

I suspect that this modification is close enough to what Picollo had in mind, and so I will move on to a more substantial worry, assuming that this mild modification of her view is in place.

Picollo shows that the Yablo paradox is not m-circular on this understanding of circularity. The argument is simple: We first introduce a predicate "S(…)" such that:[2]

S = "("∀)(y > x ∀ ~ T(S(y)))"

The Old-fashioned Yablo Paradox then consists of the ω-sequence of sentences:

S(1), S(2), S(3)…

Where:

S(1) = "("∀)(y > x ∀ ~ T(S(y)))"
S(2) = "("∀)(y > 2 ∀ ~ T(S(y)))"
S(3) = "("∀)(y > 3 ∀ ~ T(S(y)))"
Etc.

This set of sentences is (provably) non-m-circular since, as Picollo succinctly puts it, "…none of its members establishes an identity between a term and a formula that mentions it…"

The worry, however, concerns exactly what this shows. As a proof that this version of the Yablo paradox is not m-circular, it is certainly correct. And as Picollo notes, m-circularity is a substantial, non-trivial kind

[2] I am suppressing some complications, such as the need for Feferman's 'dot' notation, or something equivalent, in order to simplify the formulas.

of circularity –one that holds some promise of doing non-trivial theoretical work. But of course, we weren't originally interested in m-circularity, we were interested in circularity simpliciter. And as Piccolo herself notes, m-circularity is sufficient, but not necessary, for circularity simpliciter:

Naturally, m-reference and m-circularity entail, correspondingly, reference and circularity simpliciter, but not the other way around.

In particular, if this is right, then non-m-circularity does not entail non-circularity. So showing that the Yablo construction isn't m-circular doesn't entail that it isn't circular simpliciter. In short, Picollo's account, by her own admission, doesn't fully answer the question with which we began –whether the Yablo construction is circular– even though it makes substantial progress, insofar as it identifies an important non-trivial sort of circularity that the Yablo construction definitely does not exhibit.

So what is left? The missing part of the story is exactly what Picollo highlights immediate before making the admission about the non-necessity of m-circularity –there is a very deep, and very difficult problem regarding how we are to understand the contribution that quantifiers make to the reference of a sentence. Picollo mentions Leitgeb's (2002) attempt to sort out these difficult issues. I am not going to attempt to make any substantial contribution to that project here. Instead, I will conclude by highlighting a pair of constructions that I think will provide a useful test case for any account of 'reference' or 'aboutness' that will provide both necessary and sufficient conditions for circularity.

The two constructions in question involve the introduction of the following two predicate symbols respectively:

$S_1$ = "There are at least x natural numbers such that $\sim T(S_1(x))$"
$S_2$ = "There are at least x natural numbers greater than x such that $\sim T(S_2(x))$"

The first predicate symbol, along lines similar to the Picollo's construction of the Old-fashioned Yablo paradox, provides us the following ω-sequence of formulas:

$S_1(1)$ = "There is at least 1 natural number x such that $\sim T(S_1(x))$"
$S_1(2)$ = "There are at least 2 natural numbers x such that $\sim T(S_1(x))$"
$S_1(3)$ = "There are at least 3 natural numbers x such that $\sim T(S_1(x))$"
Etc.

The second predicate symbol provides the following ω-sequence of formulas:

$S_2(1)$ = "There is at least 1 natural number x greater than 1 such that $\sim T(S_2(x))$"
$S_2(2)$ = "There are at least 2 natural numbers x greater than 2 such that $\sim T(S_2(x))$"
$S_2(3)$ = "There are at least 3 natural numbers x greater than 3 such that $\sim T(S_2(x))$"
Etc.

Both sequences are paradoxical. But my own intuition is that the first sequence of formulas is circular in a sense in which the second sequence is not –the first, intuitively, involves quantification over all formulas in the list, while the second only involves quantification over those sentences later in the list. A complete and adequate explanation of circularity will, in my opinion, need to be able to account for differences of this sort.[3]

## 3. Response to Teijeiro

Teijeiro introduces a distinction between natural and formal paradoxes, and claims that my re-construction of the Yablo paradox in my infinitary logic $L_P$ fails to provide a formal paradox. Since there is a sense of the term 'formal paradox' –namely, a paradox formulated in some appropriate formal language– in which my reconstruction is quite obviously a formal paradox, it will be worth working out in a bit of detail exactly what Teijeiro's claim is, before assessing whether it is correct.

A *natural paradox*, according to my understanding of Teijeiro, is a paradox that arises in natural language, and whose importance stems from the fact that the (informal) principles that give rise to the paradox are ones that we have good prior reasons for wanting to accept. The Liar paradox and (perhaps) Stephen Yablo's original (1985, 1993) informal formulations of the Yablo paradox are natural paradoxes par excellence, in this sense. A *formal paradox* is a paradox that arises within a formal system, and whose importance stems from the fact that the (formal) principles that give rise to the paradox are ones that we have good reasons for wanting the formalization in question to support. Formal language versions of the Liar paradox and the Yablo paradox constructed

[3] It is worth noting that Picollo is working on just such an account of quantificational reference. While it is not the account I would have produced myself, it is sophisticated and well-motivated (and clearly superior to existing accounts in the literature on a number of fronts). I do not yet know how it will handle the two Yabloesque examples given above, but I have no doubt that the answer is interesting.

within Peano arithmetic supplemented with a Truth predicate are uncontroversial examples of formal paradoxes.

I think that this distinction, and the way that Teijeiro makes it, is of immense importance, since it emphasizes important questions regarding the connections (or lack thereof) between the natural languages we use everyday and the formal languages that logicians, philosophers, and many others routinely study. I wrote my dissertation attempting to answer questions of exactly this sort, and I worry that many of us study formal systems of various sorts without being critical enough about what exact theoretical role such constructions play. One natural thought is that formal languages are used to precisely identify, isolate, and understand problems and puzzles we find in our natural languages. In short, formal languages are models –in roughly the sense this term is used within science and its philosophy– of natural languages. And this is surely right, as a description of some uses of formal languages. But as Teijeiro emphasizes (and I agree), this does not cover all of the ways that we can, and do, use formal languages.

In particular, Teijeiro emphasizes the fact that natural paradoxes seem to involve principles that we wish to accept –that is, principles that we seem to have good reasons to think true– while formal paradoxes involve principles that we desire to hold in our formalizations. But as is clear from Teijeiro's discussion, not every principle that we believe to be (or hope to be) true in natural language need have an analogue that we believe ought to hold in relevant formal systems, nor vice versa. In fact, she points out that there can be natural paradoxes that have no formal paradox analogue, and formal paradoxes that have no natural paradox analogue. Before moving to the her main criticism of my infinitary version of the Yablo paradox, it is worth taking a closer look at these cases.

Teijeiro claims that the Sorites paradoxes are examples of natural paradoxes that have no formal paradox analogue. Although she does not explain exactly why, the reasons are not hard to reconstruct. Vagueness, and hence the Sorites paradox, involves a particular kind of indeterminacy –a lack of sharp borders– in an essential way. Any attempt to formalize the paradox in the precise mathematics presupposed by any formal language will idealize away the imprecision in question. As a result, the very act of formalization seems to eliminate the essence of vagueness itself. I have elsewhere called this problem the Problem of Inappropriate Precision (see Cook 2002, 2011).

Teijeiro also claims that the Cantor paradox is a formal paradox with no natural language analogue. One argument for this claim (I am not sure it is Teijeiro's, but it is a relatively natural thought) is that the set-theoretic paradoxes involve subtle set-theoretic principles, arguments, and concepts that were (as a matter of practice, not in principle) impossible to formulate prior to the precise formulation of set theory within a formal language. To be honest, I not sure this claim about the Cantor paradox is correct, since I am unconvinced that the Cantor paradox actually requires any concepts or principles that could not have been expressed informally in the late nineteenth century (and it seems likely that Cantor's own version of the argument was formulated in natural, not formal, language, although of course the 'mathematese' within which most real, research-level mathematics occurs likely lives in some linguistic purgatory between genuine natural languages and purely formal languages). This is of no consequence, however, since the issue is whether there is an example along these lines –the identity of the particular example matters little– and I find the argument somewhat more compelling when applied to the Burali-Forti paradox, which presupposes a rather deep and esoteric understanding of the ordinal numbers.

Even if these particular examples don't, in the end, show that the particular paradoxes in question are examples of one type of paradox with no analogue of the other type, they are sufficiently plausible to suggest the possibility of such paradoxes. In short, they make plausible the claim that not every principle we wish to accept in natural language is necessarily a principle we wish to hold in the relevant formal languages, and vice versa. As a result, the distinction between formal paradoxes and natural paradoxes is a substantial one, and does not merely reduce to the question of which language we use to formulate the paradox.

With these observations out of the way, we can address Teijeiro's main worry about my infinitary conjunction construction. Teijeiro's particular complaint about the $L_P$ reconstruction of the Yablo paradox concerns the particular reference function $\delta$ that is used to construct the problematic set of (infinitary) $L_P$ sentences, which then are shown to be deductively inconsistent in the resulting deductive system.[4] Teijeiro questions whether this construction is a genuine formal paradox by

[4] It is worth emphasizing that Teijeiro is not questioning the existence of the function in question, which can be coded up as a recursive function from natural numbers to recursive sets of natural numbers – in effect, it is the simple function:

$f(x) = \{y : y > x\}$

Rather, she is questioning whether the referential relation built upon this recursive function played or plays any substantial and necessary role in the formal language in question.

questioning whether we really have good reasons to want such a reference function within our formalization.

Now, there is one kind of reason for including such a function that we do not have. As we already noted, one (but not the only) use of formal languages is to model interesting or problematic linguistic phenomena that occur in natural language. But the present case is, arguably, not such a case: One could argue that there are no good reasons for thinking that natural language contains either infinitary conjunctions or infinitary reference relations such as $\delta$ (of course, one might argue otherwise as well –we shall return to this topic multiple times below). But I take it that Teijeiro's worry is more general than this –she sees no reason whatsoever, of any sort, for wanting such a reference function in the formal language in question.

I believe that there is such a reason, however –one that can be identified by attending to the complex flow of concepts and constructions that occurs between our natural language practices and our study of such practices via formal languages. One important aspect of this interchange is that formal constructions often help us explain, and solve, puzzles that arise with regard to those concepts and principles that play important roles in our natural language theorizing. But there is another role that formalization plays that gets less attention –the fact that formal work often shows us that natural language principles to which we have not yet paid significant attention deserve substantial scrutiny. In other words, the wider-ranging sort of formal work that goes along with the construction of, and solution to, formal paradoxes also often brings about the discovery of new paradoxes in natural languages.

I would argue that this is one of the primary roles that the infinitary $L_P$ construction plays. Due to its substitution of infinitary conjunction and infinitary referential relations for the arithmetical diagonalization methods (or, perhaps, direct stipulation) of the original Yablo paradox(es), the $L_P$ construction does not correspond to any natural paradox existing in our current finitary natural languages. But it does suggest a new natural paradox –one involving infinitary conjunction in an infinitary natural language. What is most interesting about this construction is that it might not be a genuine natural paradox for us –that is, for language users who are restricted to finitary languages and thus cannot think, or reason with, the infinitely long conjunctions involved in the $L_P$ paradox.[5] Of course, the fact that we cannot, perhaps, reason in infinitary languages does not imply

[5] In actuality I think that whether we can reason in infinitary languages is an open question(see the discussion in Cook forthcoming-a Chapter 2).

that we cannot reason about such languages. Further, if it is logically possible that there could be super-tasking infinitary beings that could reason in such natural languages, and if our philosophical account of truth should apply equally well to such languages and such language users, then we have every reason to be philosophically interested in such paradoxes in virtue of our desire to formulate a truly general account of logic and of truth. And given all this, we therefore do have good reasons for wanting our infinitary formal languages to allow for the existence of reference functions like δ. Thus, in the end, the infinitary $L_P$ construction is a genuine formal paradox after all.

## 4. Response to Pailos

Pailos identifies two places where one might object to the reconstruction of the Yablo paradox within $L_P$: One might object to the claim that it is a paradox, or one might object to the claim that the infinitary derivation is a genuine proof. Pailos concentrates on the second worry (which is convenient, since we have already devoted some discussion to Teijeiro's variant of the first worry), and argues that any reasonable account of proof that would not allow the $L_P$ derivation will result either in a theory of truth that fails to be sufficiently general, or in a theory of proof that denies that the 'Old-fashioned Liar" is paradoxical.

For the most part, I agree wholeheartedly with Pailos' conclusions, which will I suspect not surprise him since (1) he is defending my view regarding the $L_P$ reconstruction of the Yablo paradox, and (2) his own arguments are rather natural extensions of points I made towards the end of Chapter 2 of Cook (forthcoming-a) (although it should be emphasized that Pailos' takes these initial thoughts both further, and in somewhat different directions, than I do). As a result, it would be difficult indeed to criticize Pailos' main arguments. But his discussion does highlight a more general issue on which he and I differ, I think, and as a result I will focus on that issue.

In discussing my response to Priest and Beall's objections to infinitary conjunction or ω-rule versions of the Yablo paradox, Pailos writes that:

> But Cook seems to have been a little too lenient with his rivals. Let's suppose that, indeed, those limitations [to finitary methods of reasoning] are universal. (I mean those are the limitations of any rational being, and not only our particular situation.) Why should we

> base our theory of truth on those epistemic limitations? Our theory of truth is not supposed to be epistemic in any relevant sense, not even this one. There will still be inferences in those systems that preserve truth. And a theory of truth must explain why. (this journal, p. 40)

In one sense, this is certainly right. A theory about what statements are and are not true is not epistemic, and thus (assuming that the infinitary constructions are statements in the relevant sense in the first place) our theory of truth should apply to statements of all sorts, including the infinitary $L_P$ ones used in my reconstruction of the Yablo paradox, regardless of whether we, or anyone else, can actually reason with such statements.

But an account of paradoxes is not, typically, limited to an account of truth. In addition, the account will also involve an account of logic (typically, accounts modify only one of these –i.e. restrict the T-schema or opt for a non-standard logic, although there is nothing preventing theories that do both). And the question at issue does not seem to be whether the infinitary rules are truth-preserving (which they are, so long as the infinitary statements involved in application of the rules are genuine statements and thus truth-apt in the first place). The question at issue is whether the rules are logical.[6]

This brings us to the real issue –Pailos' insistence that the restriction to rules that are applicable by some possible rational being is artificial, since it imposes an illegitimate epistemic constraint on our understanding of legitimate rules of inference. But this objection presupposes a particular understanding of logic (on that, as we shall see, I don't accept).

Loosely speaking, logic is the study of the logical consequence relation –that is, it is the study of which statements follow from which other statements (or sets of statements) necessarily, and in virtue of their logical form. The expression "follows from", however, can be understood in (at least) two ways here, however. First, we have a metaphysical understanding, which can be roughly glossed as something like the following *metaphysical explication of logic*:

> *MEL*: $\Phi$ follows from $\Delta$ if and only if it is impossible that all the members of $\Delta$ are true and $\Phi$ is false (and further, this

[6] I will be the first to emphasize that I was less than clear about this distinction in my discussion of the issue in Cook (forthcoming-a).

metaphysical guarantee holds in virtue of the logical forms of the statements in $\Delta$ and or $\Phi$).

I take it that Pailos' understanding of the "follows from" relation is something along the lines of MEL. There is an epistemic alternative, however, that understands "follows from" along something more like the *epistemic explication of logic*:

$EEL_1$: $\Phi$ follows from $\Delta$ if and only if, whenever one believes all the members of $\Delta$, one ought to believe $\Phi$ (and further, this prescription holds in virtue of the logical forms of the statements in $\Delta$ and or $\Phi$).

Or perhaps somewhat more plausibly:

$EEL_2$: $\Phi$ follows from $\Delta$ if and only if, whenever one believes all the members of $\Delta$, one ought not believe any statement incompatible with $\Phi$ (and further, this proscription holds in virtue of the logical forms of the statements in $\Delta$ and or $\Phi$).

Or:

$EEL_3$: $\Phi$ follows from $\Delta$ if and only if, whenever one is justified in believing all the members of $\Delta$, then one is justified in believing $\Phi$ (and further, this epistemic guarantee holds in virtue of the logical forms of the statements in $\Delta$ and or $\Phi$).

Now, there is no doubt that, as a matter of professional practice, we tend to focus on the relation I have called MEL. This does not immediately imply that MEL explicates the relation that is of intrinsic interest to us, however. After all, as interesting as metaphysics can sometimes be, it is presumably not the abstract relation between statements that is of direct use to us, but rather the implications such relations have for what we ought to believe or what we are justified in believing.

Furthermore, if it turns out that it is one or another version of EEL that provides the correct understanding of the logical consequence relation, it is relatively easy to explain why we nevertheless pay so much attention to the (metaphysical) relations holding between statements rather than paying direct attention to the more fundamental epistemic relations. First, thanks to Frege, Tarski, et alia, our mathematical tools for studying the metaphysical relations are much more powerful and much more reliable than our tools for studying the corresponding epistemic relations. Second, even if we think the epistemic relations picked out by EEL and the metaphysical relations picked out by MEL are distinct, it is nevertheless tempting to adopt something like the following principle of covariance:

> PC: For any statement $\Phi$ and set of statements $\Delta$ where we (or some rational agents) are able to grasp $\Phi$ and all the members of $\Delta$:
>
> It is impossible that all the members of $\Delta$ are true and $\Phi$ is false if and only if whenever one is justified in believing all the members of $\Delta$, then one is justified in believing $\Phi$.

In fact, I personally believe that the best way to understand the logical consequence relation is along the lines of one or another of the variants of EEL listed above, but I think we can proceed (in most, but not all cases –see below!) as if we accepted MEL, since I accept something like PC.

Finally, however, we get to the real heart of the issue, and the connection with the infinitary conjunction version of the Yablo paradox. It is worth noting that the antecedent of PC is necessary –after all, as is clear from the discussion above, there could be cases where the metaphysical relations holds but where the epistemic relation does not in virtue of the non-graspability of the statements in question (e.g. because of their length). And, in fact, I think that there are (or at least could be) variants of the Yablo paradox that involve statements that no rational being could grasp, and whose proof of paradoxicality involves infinitary rules of inference that no rational being could grasp (and further, this limitation is an in-principle one, not one imposed by our limited biology or by contingent facts regarding the continuum-sized nature of spacetime).

Let $\Omega$ be any unbounded initial segment of the ordinals, and consider the $\Omega$-sequence of statements of the form:

$$\text{For all } \alpha \in \Omega,$$
$$S(\alpha) \leftrightarrow (\forall x > \alpha)(\sim T(<S(x)>))$$

Let us call a particular such construction the $\Omega$-*generalized Yablo sequence* (for $\Omega$ an initial segment of ordinals), and call these *generalized Yablo sequences* more generally. Any generalized Yablo sequence is paradoxical (by a simple extension of the standard Yablo reasoning), and each such paradox can be constructed within $L_P$, and proved inconsistent within the corresponding deductive system.

Now, independently of any particular views on the paradoxes, it is somewhat plausible to think that there is some ordinal $\alpha$ 'large enough' such that the $\{\beta : \beta < \alpha\}$-generalized Yablo sequence is not graspable by any (possible or actual) rational being. Note that physical possibility might disappear rather early –since there is no order-preserving embedding of an uncountable initial sequence of the ordinals into the continuum, there is at least some doubt that any rational being could grasp or manipulate physical tokens of formulas of arbitrarily infinite length. Regardless of where, exactly, the line lies, however, my own view is that there must be such a line. The reason is simple: My *Embracing Revenge* view of the Liar paradox and the Revenge phenomenon commits me to the impossibility of absolutely unrestrictedly general quantification (see Cook 2008, 2009a). In particular, we cannot quantify over all the ordinals. So the $\Omega$-generalized Yablo sequence where $\Omega$ is the proper class of all ordinals –really all of them!– is a case where no rational being, of any sort, can grasp the statements in question or carry out the reasoning to a contradiction. This does not prevent the statements in question existing in some abstract sense, perhaps, nor does it prevent the contradiction from following in the metaphysical, MEL sense of "follows from". But if, like myself, one is committed to the impossibility of absolutely general quantification, and is also committed to the primacy of an epistemic understanding of logical consequence, then there are good reasons for thinking that there is some variant of the Yablo paradox that is in principle not graspable, and not provably inconsistent, and hence the relevant proper-class-sized versions of the conjunction rules must fail to satisfy EEL.

The rhetorical move Pailos criticizes can now be defended. The argument for the existence of a genuinely non-circular variant of the Yablo paradox requires that the infinitary statements be graspable (at least, it must be logically possible that they are in principle graspable), and that the relevant applications of the infinitary inference rules codify genuine instances of logical consequence. On the metaphysical account of logic, one could argue that all the uncountable Yablo-like constructions –including the proper-class sized constructions– given above are genuine paradoxes. But given my commitment to the failure of absolutely unrestricted quantification and to the epistemic conception of logical consequence, I am forced to admit that some Yablo-like constructions of the sort given above are not genuine paradoxes. As a result, it is important to argue that although some of these infininitary, $L_P$ constructions are not genuine paradoxes, the countably infinite ones are genuine paradoxes. And this is exactly what the argument given at the end of Chapter 2 is meant to accomplish.

## 5. Response to Tajer

Tajer's response is –quite politely– formulated as an exploration of whether and how the sort of infinitary inferences used in my reconstruction of the Yablo paradox within $L_P$ might be justified within a broadly anti-realist account of the philosophy of mathematics. Of course, the considerations he brings to bear on this issue have consequences for anti-realist applications of infinitary resources in a much wider range of contexts. The politeness comes in the fact that he nowhere treats the discussion as a criticism of, or problem for, my own rather esoteric collection of philosophical commitments. As Tajer is well aware, I have long been interested in anti-realism and intuitionism (see Cook 2002, 2005), and although the first paper where I explicitly commit myself to a version of intuitionism is still forthcoming (Cook forthcoming-b), Tajer is also aware that I have been publicly self-describing as an anti-realist and intuitionist for a number of years.

As I see it, there are three separate strands of thought running through the past century of writings on intuitionism. Roughly put, these are:

(1) The status of mathematical objects, proofs, etc. as mental constructions imposes restrictions on both the logic to be applied in such contexts and on assumptions regarding infinite existents.
(2) The fact that understanding of meanings must be manifestable imposes restrictions on logic.
(3) The fact that certain concepts are indefinitely extensible imposes restrictions on assumptions regarding infinite existents (and perhaps also imposes restrictions on logic).

Strands (1) and (2) run continuously through the writings of Brouwer, Heyting, and Dummett, while (3) is a contribution due solely to Dummett.

My own anti-realist leanings are due solely to consideration (2). As a platonist and a logicist, the ontological scruples underlying (1) are not applicable to my own views (see Cook 2009b, 2012). In addition, contrary to Dummett's own position on (3), I do not find any reason to think that imposing broadly anti-realist or intuitionistic constraints on our reasoning helps us to deal with the paradoxes that arise due to indefinitely extensible concepts such as set, cardinal, ordinal, and truth value (I do think that the indefinite extensibility of these concepts impose some constraints on our logical theorizing, just not the constraints usually imposed by anti-realists (see Cook 2008, 2009a).

Thus, the question, quite rightly raised by Tajer, is whether or not manifestation concerns are compatible with the acceptance of infinitary rules of inference, such as the infinitary conjunction rules used in $L_P$. I think that Tajer is correct in suggesting that the answer is "yes", in a certain qualified sense – there is nothing in infinitary rules of inference in-and-of-themselves that violate the manifestation requirement. Anti-realism is compatible with the acceptability of infinitary inferences of the sort used in $L_P$. Nevertheless, there are some subtleties that are worth working out. The best way to show this is to compare two examples of applying the manifestation argument and its consequences to rules of inference.

In its simplest form, the manifestation argument goes something like this. Understanding an expression requires being able to manifest that understanding publicly (i.e. communicating this understanding), and this in turn amounts to demonstrating an ability to recognize when the expression does or does not apply (or, more generally, demonstrating a recognition of the truth or falsity of statements involving that expression, when such expressions are true or false, and when all other expressions that are parts of the statements in question are also understood). This leads to acceptance of the Principle of Epistemic Constraint:

PEC: For any $\Phi$, if $\Phi$ is true then $\Phi$ is in principle knowable.[7]

Tajer does a good job of surveying the subtleties of what, exactly, we mean by 'in principle', so I won't tarry there (nor will I address other well-known issues plaguing this principle, such as the Fitch paradox). Instead, let us look at an example where PEC is used to argue that a particular logical law is invalid, and then determine whether similar worries might plague infinitary rules of inference such as those used in $L_P$.

The standard example to trot out in discussions of PEC is the *law of excluded middle*, but since we are interested in rules of inference and not in axioms, the rule of *classical reductio* will serve us better. The rule of classical reductio states that, if one assumes $\sim \Phi$ and then deduces a contradiction, then one can conclude $\Phi$ (and discharge the assumption that

[7] Note that, contrary to what is often claimed by both intuitionists and their critics, acceptance of this principle –or of intuitionism more generally– requires neither abandoning the notion of truth in favor of knowledge, nor identifying truth and knowledge. PEC requires merely that truth and knowledge co-vary, and is compatible with these nevertheless being distinct concepts.

$\sim \Phi$). Recalling that, from the PEC perspective, $\sim \Phi$ is true if and only if $\sim \Phi$ is provable if and only if $\Phi$ is refutable, our sub-proof that $\sim \Phi$ entails a contradiction shows (in effect) merely that it cannot be the case that $\Phi$ is refutable (or, equivalently, that we can refute the claim that $\Phi$ is refutable). To conclude from this evidence that $\Phi$ must thereby be true –and hence, by PEC, provable– amounts to assuming that every sentence is either refutable or true (and hence either refutable or provable). But it is exactly this sort of epistemic determinacy that the intuitionist wishes to avoid (and should avoid, given the plausible thought that logic plus the sort of meaning-theoretic considerations involved in the manifestation argument should not, alone, entail the decidability of every sentence in our language, even if our language were to turn out to be decidable in the relevant sense for some other reason). Hence, for the anti-realist, classical reductio is invalid.

But what about infinitary conjunction rules? First off, it is worth noting that the *infinitary conjunction elimination rule*:

$$\frac{\Phi_1 \wedge \Phi_2 \wedge \Phi_3 \wedge \ldots \wedge \Phi_{n-1} \wedge \Phi_n \wedge \Phi_{n+1} \wedge \ldots}{\Phi_k}$$

is, quite definitely, anti-realistically acceptable. If we know an infinite conjunction, then knowledge of any particular conjunct is presumably unproblematic and straightforward. Of course, there remains the question of whether we could ever apply this rule, because there remains the question of whether we can ever be a position where we have knowledge of an infinite conjunction. And this quite naturally brings us to the hard case –the *infinitary conjunction introduction rule*:

$$\frac{\begin{array}{l}\Phi_1\\ \Phi_2\\ \Phi_3\\ \vdots\\ \Phi_{n-1}\\ \Phi_n\\ \Phi_{n+1}\\ \vdots\end{array}}{\Phi_1 \wedge \Phi_2 \wedge \Phi_3 \wedge \ldots \wedge \Phi_{n-1} \wedge \Phi_n \wedge \Phi_{n+1} \wedge \ldots}$$

Determining the status of this rule is much more difficult.

On the one hand, this rule does not seem to be susceptible to the immediate sort of objection leveled against classical reductio above.

Assuming that the in-principle knowability of infinitely many conjuncts entails the in-principle knowability of the relevant infinite conjunction does not seem to imply implausible or unacceptable epistemic principles in the same manner that classical reductio entails global decidability (and this is not surprising, from a technical perspective, since adding infinitary connectives to intuitionistic logic does not make the logic classical in the same manner as adding classical reductio does).

At this point a comparison with the finitary conjunction introduction rules is helpful. The reason that $\Phi$ and $\Psi$ individually entail $\Phi \wedge \Psi$ is that we are assured that, given a proof or other justification for $\Phi$ and another proof or justification for $\Psi$, we can combine these to arrive at a single, composite proof or justification of $\Phi \wedge \Psi$. The question then becomes this: Is it always possible to combine infinitely many justifications into a single justification in the relevant sense, turning a many into a one, so to speak?

As Tajer notes, I have argued that this is, at least in principle, physically possible for supertasks. But this only gets us countably infinite conjunctions. Tajer argues convincingly that, since any more stringent means for drawing the relevant line –such as finitism– must already idealize past what is actually humanly possible, there is no principled reason to rule out infinite constructions as constituting genuine knowledge (or at least its possibility) in the relevant sense. This brings us back to questions familiar from the discussion of Pailos' contribution above: How big can the conjunctions be? And the answer will again depend on what one thinks the relevant criteria are. Given my commitments already outlined above, it is clear that proper-class-sized conjunctions are too large, but countably infinite conjunctions are okay.

As a result, we will likely find ourselves with a mixed view of the sort hinted at in the end of Tajer's contribution. For some small cardinals $\kappa$, it might be the case that the conjunction introduction rule for conjunctions of length $\kappa$ is logically valid (i.e. always acceptable as a matter of logical form) in virtue of the fact that passing from knowledge of the individual conjuncts to knowledge of the conjunction is always logically possible. And for some large cardinals $\kappa$ (as well as for conjunctions of proper-class size), it might be the case that the conjunction introduction rule for conjunctions of length $\kappa$ is never logically valid in virtue of the fact that conjunctions of that size are always impossible to grasp 'all at once', so to speak. But in the middle, the may well be cardinals $\kappa$ such that we can legitimately apply the introduction rule to some conjunctions of length $\kappa$ but not to others, in virtue of the fact that some conjunctions of this length (and the individual justifications for their infinitely many conjuncts) can be grasped 'all at once', but other conjunctions of this length (and their justifications) cannot be so grasped. Note that, if logic is formal, then applications of these instances of the rule are never logically valid (since formality implies that either all instances are valid or none are), but the point is that some of these instances might still be truth preserving and graspable, and hence can be legitimately applied when reasoning. When Tajer writes that:

> If we relax the requirements of manifestation and decidability in principle, it results that our use of rules like those in some occasions (for example, when we know that they preserve truth) doesn't alter any fundamental principles of anti-realism. (this journal, p. 50)

It sounds very much like he has this middle class in mind.

## 6. Response to Rosenblatt

Rosenblatt focuses on the third chapter of the book, and in particular where I consider Sorensen's General Purge of Self-reference. He argues that we might be able to carry out a general version of the purge in a richer, infinitary language much like $L_P$, but containing modal and epistemic notions such as possibility, knowability (or knowledge simpliciter), and believability (or belief). Rosenblatt surveys a number of possible such constructions, and convincingly demonstrates that there is potential here for handling a wide variety of paradoxes within such an enriched system. One interesting aspect of the project that Rosenblatt mentions by does not work out in detail is a Yabloesque puzzle that I also ignore in my book – a Yabloesque version of the paradox of believability (or belief).[8]

First, I should mention that much of what follows is due to joint work with Cat Saint Croix during an independent research project at the University of Minnesota.

We will first consider the paradox in question within arithmetic. Consider the result of extending Peano Arithmetic with an idealized 'rational belief' predicate B(…) – i.e. a predicate that holds of the Gödel code of a sentence if and only if one has good reasons (or warrant) for believing the sentence. The following principles plausibly hold for the rational belief predicate B(…):

[8] In what follows I am going to construct the paradox in terms of an idealized 'rational believability' predicate. The paradox could, of course, be reformulated in terms of a 'rationally believed' predicate, along lines similar to those for transforming the 'is knowable' version of the Montague paradox into the 'is known' version.

*Necessitation*: If $\Phi$ is a theorem, then $B(<\Phi>)$ is a theorem.
*Closure*: All instances of:

$$B(<\Phi \rightarrow \Psi>) \rightarrow (B(<\Phi>) \rightarrow B(<\Psi>))$$

are theorems.

Note that we obtain no paradox from these principles alone –the Gödelian provability predicate satisfies analogues of both Necessitation and Closure. Further, as Rosenblatt notes, the factivity principle for belief:

*Factivity*: All instances of $B(<\Phi>) \rightarrow \Phi$ are theorems.

Is rather implausible. But we can obtain a paradox with the addition of a more plausible principle –the *Principle of Transparency of Disbelief*:

TransDisb: All instances of:

$$B(<\sim B(<\Phi>)>) \longrightarrow \sim B(<\Phi>)$$

Are theorems.

(Rosenblatt calls this principle –somewhat more accurately– B-semifact). Given this principle, we can derive a contradiction. We need only consider the sentence:

$$\Phi \leftrightarrow \sim B(<\Phi>)$$

Obtained via diagonalization.

Proof of contradiction:

| | | |
|---|---|---|
| [1] | $\Phi \leftrightarrow \sim B(<\Phi>)$ | Diagonalization. |
| [2] | $B(<\Phi>) \leftrightarrow B(<\sim B(<\Phi>)>)$ | 1, Closure. |
| [3] | $\quad B(<\Phi>)$ | Assumption. |
| [4] | $\quad B(<\sim B(<\Phi>)>)$ | 2, 3, Logic. |
| [5] | $\quad \sim B(<\Phi>)$ | 4, TransDisb. |
| [6] | $\quad \perp$ | 3, 5, Logic. |
| [7] | $\sim B(<\Phi>)$ | 3 – 6, Reductio. |
| [8] | $\Phi$ | 1, 7, Logic. |
| [9] | $B(<\Phi>)$ | 8, Necessitation. |
| [10] | $\perp$ | 7, 9, Logic. |

Interestingly, however, if we unwind this construction according to the recipe given in Cook (forthcoming-a), we don't seem to get a paradox –at least, not solely using the principles listed above. Applying the unwinding, we obtain:

$$(\forall x)(\Phi(x) \leftrightarrow (\forall y)(y > x \longrightarrow {\sim}B(<\Phi(y)>)))$$

Consider the following ‘proof’:

| | | |
|---|---|---|
| [1] | $(\forall x)(\Phi(x) \leftrightarrow (\forall y)(y > x \longrightarrow {\sim}B(<\Phi(y)>)))$ | Diagonalization. |
| [2] | $B(<\Phi(a)>)$ | Assumption. |
| [3] | $\Phi(a) \leftrightarrow (\forall y)(y > a \longrightarrow {\sim}B(<\Phi(y)>))$ | 1, Logic. |
| [4] | $B(<\Phi(a)>) \leftrightarrow B(<(\forall y)(y > a \longrightarrow {\sim}B(<\Phi(y)>))>)$ | 3, Closure. |
| [5] | $B(<(\forall y)(y > a \longrightarrow {\sim}B(<\Phi(y)>))>)$ | 2, 4, Logic. |
| [6] | $B(<(\forall y)(y > a+1 \longrightarrow {\sim}B(<\Phi(y)>))>)$ | 5, Closure. |
| [7] | $\Phi(a+1) \leftrightarrow (\forall y)(y > a+1 \longrightarrow {\sim}B(<\Phi(y)>))$ | 1, Logic. |
| [8] | $B(<\Phi(a+1)>) \leftrightarrow B(<(\forall y)(y > a+1 \longrightarrow {\sim}B(<\Phi(y)>))>)$ | 7, Closure. |
| [9] | $B(<\Phi(a+1)>)$ | 6, 8, Logic. |
| [10] | $(\forall y)(B(<y > a>) \longrightarrow B(<{\sim}B(<\Phi(y)>)))$ | 5, Closure, ??? |
| [11] | $B(<a+1 > a>) \longrightarrow B(<{\sim}B(<\Phi(a+1)>))$ | 10, Logic |
| [12] | $a+1 > a$ | Arithmetic. |
| [13] | $B(<a+1 > a>)$ | 12, Closure. |
| [14] | $B(<{\sim}B(<\Phi(a+1)>))$ | 11, 13, Logic. |
| [15] | ${\sim}B(<\Phi(a+1)>)$ | 14, TransDisb. |
| [16] | $\perp$ | 9, 15. Logic. |
| [17] | ${\sim}B(<\Phi(a)>)$ | 2 – 16, Reductio. |
| [18] | $(\forall x)({\sim}B(<\Phi(x)>))$ | 17, Logic. |
| [19] | ${\sim}B(<\Phi(0)>)$ | 18, Logic |
| [20] | $(\forall y)(y > 0 \longrightarrow {\sim}B(<\Phi(y)>))$ | 18, Logic. |
| [21] | $\Phi(0)$ | 1, 21, Logic. |
| [22] | $B(<\Phi(0)>)$ | 21, Necessitation. |
| [23] | $\perp$ | 19, 22, Logic. |

This does constitute a proof of a contradiction from the universally quantified generalization, just as we obtained in the unwinding of the Knower paradox and in the Unwinding of the Liar paradox (the Yablo paradox). But there is an additional principle that we needed to invoke –one not needed in Knower or Yablo case. The proof requires us to move from:

$$B(<(\forall y)(y > a \longrightarrow {\sim}B(<\Phi(y)>))>)$$

At line [5] to:

$$(\forall y)(B(<y > a>) \longrightarrow B(<\sim B(<\Phi(y)>)))$$

at line [10]. Ignoring applications of closure, this amounts to assuming a version of the converse Barcan formula for the believability predicate:

$$B(<(\forall x)(\Phi(x))>) \longrightarrow (\forall x)(B(< (\Phi(x))>)$$

Thus, unlike the case of the Liar and the Knower, the unwinding of the paradox of transparent disbelief seems to require an additional principle that was not needed in the single-sentence version.

Now, I carried out the above derivations within arithmetic, taking advantage of quantifiers, instead of within an infinitary system such as $L_P$. So, given that Rosenblatt is concerned with whether or not we can carry out the general purge of self-reference within such infinitary systems, one might rightly wonder at this point what one has to do with the other. The answer is simple: A similar phenomenon will occur within any attempt to derive a contradiction from an infinite conjunction version of the paradox of transparent disbelief. In order to carry out an infinitary derivation modeled on the arithmetic one just given, one will need to assume an additional rule of the form:

$$\frac{B(\Phi_1 \wedge \Phi_2 \wedge \Phi_3 \wedge \ldots \wedge \Phi_{n-1} \wedge \Phi_n \wedge \Phi_{n+1} \wedge \ldots)}{B(\Phi_1) \wedge B(\Phi_2) \wedge B(\Phi_3) \wedge \ldots \wedge B(\Phi_{n-1}) \wedge B(\Phi_n) \wedge B(\Phi_{n+1}) \wedge \ldots}$$

(or some other rule equivalent to, or implying, this rule). The point is not that this rule is implausible –on the contrary, it is a very natural generalization of the intuition underlying the finitary closure rule. The point, simply put, is that it is a new rule –one not required in the proof of a contradiction from the original single-sentence paradox of transparent disbelief. Further, it is a rule that is specific to the believability predicate –all of the infinitary rules introduced in $L_P$ to deal with the Yablo paradox and the unwinding of the Knower paradox were rules for the logical connectives, not rules for the problematic, paradox-prone predicates. In short, carrying out the proof given above of a contradiction for the unwinding of the paradox of transparent disbelief seems to require us to add, not only new rules for the infinitary logical connnectives, but new rules for the believability predicate itself –rules that were not required for the proof of contradiction in the finitary case. This seems against the spirit, if not the letter, of the general purge.

As a result, the unwinding of the paradox of transparent disbelief seems to have a very different character from the unwindings of the other paradoxes we have looked at. This seems to be an additional stumbling block in the attempt to formulate a single context, and single system, within which we can carry out unwindings for all sentential paradoxes.[9]

## 7. Response to Ojea

I will begin with Ojea's observation that my proof that the $L_P$ version of the Yablo paradox is insufficient. In short, Ojea is absolutely right. My definition of fixed-point:

> $\Phi$ is a weak sentential fixed point of $<\{S_\beta\}_{\beta\in B}, \delta>$ if and only if there is an $\alpha \in B$, a formula $\Psi$ and statement name $S_\gamma$ occurring in $\Psi$ such that:
>
> $$D(S_\alpha) = \Phi$$

And both:

$$\Phi \Rightarrow \Psi[S_\gamma/S_\alpha]$$
$$\Psi[S_\gamma/S_\alpha] \Rightarrow \Phi$$

only covers self-reference in the literal sense, not circular 'loops' more generally. As a result, he is right to note that his three-sentence loop is just as free from weak sentential fixed points as is the $L_P$ version of the Yablo paradox. The problem can be straightforwardly rectified, however, by adopting the following more general notion:

> $\Phi_1, \Phi_2,\ldots \Phi_n$ is a weak sentential cycle (of length n) of $<\{S_\beta\}_{\beta\in B}, \delta>$ if and only if there is are $\alpha_1, \alpha_2,\ldots\alpha_n \in B$, formulas $\Psi_1, \Psi_2,\ldots\Psi_n$ and

[9] Since presenting these comments, Rosenblatt has discovered an alternate proof of a contradiction (both for the arithmetic unwinding of the paradox of transparent disbelief and its infinitary variant) that does not require the additional rule. I have decided to leave my comments as they are, however, since (1) they still suggest an interesting open question regarding whether there exist any single-sentence paradoxes whose unwindings require a converse Barcan-style formula in order to prove inconsistency, and (2) there remain interesting unanswered questions regarding why some proofs of a contradiction from a single-sentence paradox can be 'mimicked' in the infinitary unwinding case without the use of the converse Barcan-style formula (e.g. the derivation of the paradox of disbelief Rosenblatt used to find his infinitary proof) and other proofs for the single sentence case cannot be generalized without the converse Barcan-style formula (e.g. the derivation I give above).

statement name $S_{g1}$, $S_{g2,}$… $S_{gn}$ occurring, respectively, in $\Psi_1$, $\Psi_2$,… $\Psi_n$ such that:

$$D(S_\kappa) = \Phi_k$$

And all of:

$$\begin{aligned}
&\Phi_1 \Rightarrow \Psi_2[S_{g2}/S_2]\\
&\Psi_2[S_{g2}/S_2] \Rightarrow \Phi_1\\
&\Phi_2 \Rightarrow \Psi_3[S_{g3}/S_3]\\
&\Psi_3[S_{g3}/S_3] \Rightarrow \Phi_2\\
&\vdots \qquad \vdots \qquad \vdots\\
&\Phi_n \Rightarrow \Psi_1[S_{g1}/S_1]\\
&\Psi_1[S_{g1}/S_1] \Rightarrow \Psi_n
\end{aligned}$$

The proof that there are no weak fixed points in the Yablo construction can be generalized to a proof that there are no weak sentential cycles in that construction (the proof is a bit lengthy, and will not be included here, but will be added to the final draft of Cook forthcoming-a). Obviously, however, Ojea's three-sentence construction showing the inadequacy of the fixed-point notion is a weak sentential cycle.

Ojea then goes on to consider a way of showing that the $L_P$ construction is non-circular – a version of the structural collapse account that proceeds by (i) attempting to defuse my own criticisms of this account, and (ii) arguing that a particular version of non-well founded set theory gets things 'right'. Ojea's central claim against my two objections is that:

> … both arguments assume that the structural collapse approach fails in mapping sets of statements into (non-well-founded) sets in a way that the referential structure of the statements is isomorphic to the membership relation of the sets. (this volume, p. 63)

I think that Ojea is right that this sums up most of the first worry. But the second worry is, I think, deeper. Even if the mapping gets the referential relations correct, there is still a question regarding whether there is some appropriately 'bad' property shared by all and only those sets that are the image of paradoxical sets of sentences, such that the 'badness' of the sets can be used to explain the 'badness' of the paradoxical sentences. As Ojea notes, the 'paradoxical' sets will fail to have colorings of a certain sort, but whether or not sets can be colored in certain way seems to be a purely formal property –not one that connects to any sort of 'badness' that can be used to explain and analyze the 'badness' of the corresponding sets of sentences and their referential relations. Thus, this worry wasn't one about getting the referential relations right, but rather concerned the explanatory role that such a mapping could play even if we do get the referential relations correct.

Ojea goes on to consider the four well-known variants of the anti-foundation axiom, and argues that it is only within BAFA that the mapping from sets of sentences to non-well-founded sets gets the referential relations right. There are two worries about this argument.

The first concerns the structure of the argument, or, slightly better put, what would need to be done to formulate a complete proof that BAFA gets the referential relations right. Ojea's argument is inductive (in the scientific sense of the term) –he considers a handful of examples, and notes that only BAFA, and not FAFA, SAFA, or AFA, gets the reference relations correct. If the only options were these four, then this would be enough to settle matter. But these four set theories were formulated with set-theoretic goals in mind. Hence it is possible that they are the most interesting, most fruitful, or most powerful non-well-founded set theories when viewed from the perspective of pure set theory, but that some other set theory (presumably between BAFA and FAFA) is the right set theory when viewed from the perspective of mapping sets of $L_P$ statements to non-well-founded sets. In short, Ojea's arguments are not sufficient to rule out some set theory intermediate between BAFA and FAFA as the right one for the job.

How to rectify this? The obvious approach would be to abandon the case-by-case methodology, and attempt to prove inductively (in the mathematical sense of induction) that the mapping based on BAFA gets the referential relations correct for every set of $L_P$ sentences. There are two problems with carrying this out, however. First, Ojea's understanding of 'getting the referential relations correct' involves, at the very least, that referential patterns that differ with regard to the paradoxical/determinate/indeterminate trichotomy get mapped to distinct non-well-founded sets This is the core of his existing argument for BAFA, since the other three set theories fail in this regard since they map the open pair:

$S_1$: $S_2$ is false.
$S_2$: $S_1$ is false.

To the same set as the Liar paradox. But, as noted in the first chapter of Cook (forthcoming-a), the paradoxicality/non-paradoxicality question for constructions within $L_P$ is in-principle undecidable, and even settling this question for the finite sets of statements is equivalent to an outstanding open question in graph theory. As a result, the likelihood of actually proving that some mapping gets the referential relations right for all $L_P$ constructions seems low.

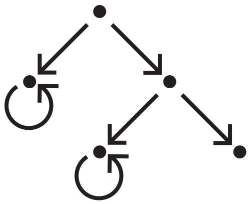
Second, and perhaps more importantly, there is a vicious circularity threatening this whole endeavor. In order to carry out a genuine proof that BAFA really does get the referential relations right, we would need to verify that the appropriate connections obtain between the referential relations of the Yablo paradox and the membership relations holding in the set BAFA maps it onto. But this requires knowing what the referential relations of the Yablo paradox are. So the argument that BAFA is the correct set theory to use in the structural collapse account seems to require that we already know what the referential relations are, defeating the entire purpose of introducing BAFA in the first place (BAFA was supposed to help answer questions about referential relations!) In short, proving that BAFA is the right set theory in this context seems hopelessly question-begging.

Nevertheless, even if the case-by-case, inductive argument is the best we can do, it does present a rather compelling case for BAFA. And Ojea has successfully shown that, of the four widely known options for axiomatizing non-well-founded sets, only BAFA will do the job in the present context. So let's grant Ojea's main claim –that if any of these theories captures the referential relations of $L_P$ sentences correctly, it is BAFA. Is this enough to salvage the structural collapse account?

Not quite. There is one last worry that needs to be addressed. The underlying intuition behind the structural collapse account is that it maps sets of statements onto (non-well-founded) sets such that the referential relations of the sentences are isomorphic to the set-theoretic membership relation on the sets in question. The thought is that we can settle issues regarding referential circularity by mapping such constructions to a context where we do understand circularity –set theory and its membership relation. But there are good reasons for thinking that BAFA is not a theory of sets.[10]

[10] Of course, the mathematical theory obtained by adding BAFA to the axioms of set theory (minus foundation) is a perfectly legitimate mathematical theory –one that describes a perfectly real mathematical structure, one that can be and has been studied just like any other mathematical structure. The point is that the structure described by BAFA does not consist of sets, but of some other (similar, but distinct) kind of mathematical object.

Adam Rieger has argued that it is not BAFA, but FAFA that provides the right account of non-well-founded sets. The argument is simple – even though all of AFA, SAFA, FAFA, and BAFA obey the letter of the axiom of extensionality, BAFA violates its spirit. To see why, consider the following graph (a modification of one of Rieger's examples):



This graph is, according to BAFA, an exact picture. Note, however, that this graph has features that seem to violate basic intuitions regarding sets. Let us call the set pictured by this graph A. Note that A contains two distinct Quine atoms B (on the left) and C (in the center) corresponding to the two terminal 'loops' (A Quine atom is a set Θ such that Θ = {Θ}). As a result, the graph above is actually an exact picture of two distinct sets. One of those sets, the one we have decided to call A, has as its members B and the set {C, ∅}, while the other has as its members C and {B, ∅}. But these two (pure!) sets, as a result of having the same graph, have exactly the same membership relation structure. While this does not violate the letter of the extensionality axiom, it seems to badly violate its spirit.

As a result, there seem to be good reasons for thinking that (at least some of) the objects described by BAFA aren't sets (even if they are very similar, in many formal respects, to sets). As a result, if we map $L_P$ constructions into the objects provided by BAFA, then we are not necessarily mapping circular and non-circular reference relations onto circular and non-circular sets. But wasn't much of the point of the structural collapse account to draw tight, theoretically useful connections between sentences (and their referential behavior) and sets (and the membership relation)?

## 8. Concluding Remarks

I have thoroughly enjoyed, and deeply benefitted from, the exchange of ideas that occurred during my two visits to Argentina –an exchange of ideas that can only be hinted at in the brief space allotted here. I would like to close this by pointing out that the enjoyment, and benefits, which came with this experience (and I hope they were mutual) went deeper than the typical questions/criticisms/responses sort of back-and-forth which might be typical of such a set-up. In addition, my interactions with the logic students and faculty in Buenos Aires have forced me to think about my own views in new and important ways.

I have occasionally joked in the past that making one's various philosophical views consistent with one another was a task best left for old age –as a young philosopher I was content to just develop the individual views. Although I don't know if I ever truly believed this claim, it is likely that at some level I hoped it was true, since the collection of views that I have defended or been at least sympathetic towards –intuitionism, neo-logicism, logical pluralism, many-valued logics for truth, degree-theoretic semantics for vagueness, etc.– seem difficult to square with one another. Apparently, however, squaring is a task for middle age –forty years, in fact (not so old, I hope). I achieved this notable age only a few weeks before arriving in Argentina for the second time. As the reader will have noticed, upon arrival I was immediately besieged with the challenges, criticisms, and comments included above –challenges, criticisms, and comments that forced me to consider the connections between my work on the Yablo paradox and my view on most of these other topics. As a result, the exchange of ideas recorded above has not only given me an opportunity to discuss and defend my thoughts on a particular topic, but it has given me an opportunity to think much more deeply, and much more critically, about all of my views. As a result, I feel like I understand my own beliefs, and my own philosophical methodology for reaching such beliefs, much better than I did mere months ago. And for that I am deeply grateful.

## References

Cogburn, J. and Cook, R. (2000), "What Negation is Not: Intuitionism and '0 = 1'", *Analysis,* 60, pp. 5-12.

Cook, R. (2002), "Vagueness and Mathematical Precision", *Mind,* 111, pp. 225-247.

——— (2005),"Intuitionism Reconsidered", in Shapiro, S. (ed.), *The Oxford Handbook of the Philosophy of Mathematics and Logic*, Oxford, Oxford University Press, pp. 387-411.

——— (2008), "Embracing Revenge: On the Indefinite Extensibility of Language", in Beall, JC. (ed.), *Revenge of the Liar*, Oxford, Oxford University Press, pp. 31-52.

——— (2009a), "What is a Truth Value, and How Many Are There?", *Studia Logica,* 92, pp. 183-201.

——— (2009b) "New Waves on an Old Beach: Fregean Philosophy of Mathematics Today", in Linnebo, O. and Bueno, O. (eds.) (2009), *New Waves in Philosophy of Mathematics*, Surrey, UK, Ashgate, pp. 13-34.

Cook, R. (2011) "Vagueness and Meaning Theories", in Ronzitti, G. (ed.), *Vagueness: A Guide, Logic, Epistemology and the Unity of Science,* Vol 19, Dordrecht, Springer, pp. 83-106.

—— (2012), "Conservativeness, Stability, and Abstraction", *British Journal of Philosophy of Science,* 63, pp. 673-696.

—— (forthcoming-a), *The Yablo Paradox: An Essay on Circularity*, Oxford, Oxford University Press.

—— (forthcoming-b) "Should Antirealists be Antirealists About Antirealism?", *Erkenntnis.*

Rieger, A. (2000), "An Argument for Finsler-Aczel Set Theory", *Mind*, 109, pp. 241-253.