ON THE POSSIBILITY OF A GENERAL PURGE OF SELF-REFERENCE

LUCAS ROSENBLATT
Universidad de Buenos Aires – CONICET
l_rosenblatt@hotmail.com

Abstract

My aim in this paper is to gather some evident in favor of the view that a general purge of self-reference is possible. I do this by considering a modal-epistemic version of the Liar Paradox introduced by Roy Cook. Using yabloesque techniques, I show that it is possible to transform this circular paradoxical construction (and other constructions as well) into an infinitary construction lacking any sort of circularity. Moreover, contrary to Cook's approach, I think that this can be done without using any controversial multi-modal rules, i.e., the usual rules from normal epistemic and modal logic are enough to show the paradoxicality of the infinitary construction.

KEY WORDS**:** Yablo's Paradox; Self-reference; Infinitary Logic; Modal Logic.

Resumen

Mi objetivo en este trabajo es ofrecer cierta evidencia a favor de la tesis según la cual una purga general de la autorreferencia es posible. Hago esto considerando una versión modal-epistémica de la Paradoja del Mentiroso introducida por Roy Cook. Usando técnicas yablescas, muestro que es posible transformar esta construcción paradójica circular (y también otras construcciones) en una construcción infinitaria que carece de cualquier forma de circularidad. Más aún, en contra de la propuesta de Cook, muestro que esto puede hacerse sin utilizar ninguna regla multimodal controversial, esto es, las reglas usuales de la lógica modal y la lógica epistémica son suficientes para mostrar la paradojicidad de la construcción infinitaria.

PALABRAS CLAVE**:** Paradoja de Yablo; Autorreferencia; Lógica infinitaria; Lógica modal.

Some philosophers[1] have seen in Yablo's paradox the starting point of a general method for purging every self-referential paradox of its self-referentiality[2]. Sorensen, for example, says that "[t]he simplicity of Yablo's paradox invites the conjecture that all [of the self-referential] paradoxes can be purged of self-reference" (1998, p. 150). If such a method were available, there would be a compelling argument in favor of the view

[1] In particular, Sorensen (1998) and Schlenker (2007).
[2] I will use 'self-referential` and ´circular` interchangeably throughout the paper.

that paradoxicality must not (maybe never) be blamed on circularity. Cook (2013) rightly notices that there are two ways to make Sorensen's idea more precise:

Weak Purge: Given any self-referential construction Σ in some language L, there is some (possibly distinct) language L* such that L* contains a non-self-referential Yabloesque analogue of Σ.

Strong Purge: Given any self-referential construction Σ in some language L, there is, in L itself, a non-self-referential Yabloesque analogue of Σ.

In his book Cook presents a purposed-build infinitary language L_P in which certain paradoxes can be represented. This language only contains one type of formula: (possibly infinite) conjunctions of sentences of the form 'S_n is false', where S_n is the name of a sentence of L_P. Cook uses a function δ to provide the denotation of every sentence name. For example, the Liar sentence can by expressed by $F(S_1)$, where $\delta(S_1) = F(S_1)$. And the Yablo sequence can be expressed as "the unwinding" of the Liar sentence[3]:

$$\delta(S_{<1,1>}) = F(S_{<2,1>}) \wedge F(S_{<3,1>}) \wedge F(S_{<4,1>}) \wedge \ldots$$
$$\delta(S_{<2,1>}) = F(S_{<3,1>}) \wedge F(S_{<4,1>}) \wedge F(S_{<5,1>}) \wedge \ldots$$
$$\delta(S_{<3,1>}) = F(S_{<4,1>}) \wedge F(S_{<5,1>}) \wedge F(S_{<6,1>}) \wedge \ldots$$
$$\ldots$$

Cook thinks that if the strong version of the purge is to have any chance of success, the language in which to carry out the unwindings must be at least as strong as the language of arithmetic. This gives us a sort of dilemma. On the one hand, since L_P is specifically designed to model certain paradoxes, it can only be used to argue in favor of the weak version of the purge. On the other hand, if the language of arithmetic is used, then it can be shown that Yablo's sequence is circular (provided that circularity is understood as being a fixed point (of a certain sort))[4]. Cook's conclusion is that unwindings are not useful to carry out the strong version of the purge.

[3] The pairs are ordered by the usual lexicographical order. The reason for using pairs (or n-tuples, depending on the example) in the unwinding has to do with the possibility of unwinding a sequence of sentences. The reader interested in the technical details for obtaining the unwinding of an L_P sentence (or sequence of sentences) can see Cook (forthcoming).

[4] Cook (forthcoming) has put forward a couple of important distinctions concerning the different types of fixed points (strong/weak, sentential/predicative). Since I do not think (see below) that fixed points provide an adequate definition of circularity, I will not concern myself with these subtleties here.

I will claim that this is a false dilemma. Even if L_P is not a good candidate to carry out the purge, it is possible to use a richer infinitary language (or maybe even arithmetic itself[5]) to argue in favor of the strong version of the purge. Of course, the success of the purge will depend on our ability to provide a definition of circularity suitable for the language. Presenting such a definition is beyond the scope of this paper, but it is interesting to note that Cook does not consider the possibility of vindicating the strong version of the purge not by finding some method different from the one based on unwindings but by finding a different definition of circularity.

With this conceptual background in mind, let me illustrate how the same unwinding operation performed on the Liar can be performed on non-semantic circular constructions. To do things properly I am going to work with an infinitary propositional language $L_{\Box KT(x)}$[6]. The vocabulary of $L_{\Box KT(x)}$ is the following: a class C of sentence names $\{S_0, S_1, S_2,.....\}$, a set of contingent sentences $\{p_0, p_1, p_2,...\}$[7], *falsum* ($\bot$), negation ($\neg$), conjunction ($\wedge$), the material conditional ($\supset$), the truth predicate (T), the possibility operator ($\Diamond$), and the knowledge operator (K). There is a denotation function δ: C $\rightarrow$ {sentences of $L_{\Box KT(x)}$} that assigns a semantic value (in particular, a sentence of $L_{\Box KT(x)}$) to each of the sentence names in C.

The non-semantic example Cook mentions is the Modal Knower[8], the sentence that says of itself that it is not knowable. In my present framework this can be formalized as $\delta(S_1) = \neg\Diamond KT(S_1)$

The unwinding of this sentence is:

$\delta(S_{<1,1>}) = \neg\Diamond KT(S_{<2,1>}) \wedge \neg\Diamond KT(S_{<3,1>}) \wedge \neg\Diamond KT(S_{<4,1>}) \wedge ....$
$\delta(S_{<2,1>}) = \neg\Diamond KT(S_{<3,1>}) \wedge \neg\Diamond KT(S_{<4,1>}) \wedge \neg\Diamond KT(S_{<5,1>}) \wedge ....$
$\delta(S_{<3,1>}) = \neg\Diamond KT(S_{<4,1>}) \wedge \neg\Diamond KT(S_{<5,1>}) \wedge \neg\Diamond KT(S_{<6,1>}) \wedge ....$
....

[5] It would have to be second-order arithmetic, since Yablo's sequence is satisfiable in first-order arithmetic.

[6] I assume that $L_{\Box KT(x)}$ does not count as a purposed-build language. It is just an extension of some infinitary language $L(\kappa,\omega)$.

[7] We need contingent sentences because otherwise truth and necessity will collapse.

[8] Actually Cook talks about the Knower, but he formalizes it as $\delta(S_1) = UK(S_1)$, and says that 'UK(x)' should be read as 'x is not knowable'. So it appears that he is thinking about a modal version of the Knower. Even if that is not the case, the modal version is interesting in its own right. (The reason for using a version without negation is that the proof showing that the unwinding of the Modal Knower does not contain fixed points depends on the language failing to contain a primitive negation operator. However, since I will not give an account of circularity in terms of fixed points, negation will be introduced as a primitive symbol).

To prove the paradoxicality of these constructions Cook suggests using the following two multi-modal rules[9]:

Rule K1: if $T(S_m) \vdash T(S_n)$ and $T(S_m) \vdash \neg\Diamond KT(S_n)$, then $\vdash \neg\Diamond KT(S_m)$
Rule K2: if $T(S_m) \vdash \bot$, then $\vdash \neg\Diamond KT(S_m)$

Another important aspect of the system is that given a particular denotation function δ, we can extend the deductive system by incorporating rules for it:

δ-*elim* if $\delta(S_n)$, then infer $T(S_n)$
δ-*intro* if $T(S_n)$, then infer $\delta(S_n)$

So the consistency of the extended system will depend on the denotation function being used. Interestingly, these rules are enough to obtain a contradiction from the Modal Knower and from its unwinding.

Another issue is whether the same sort of argument is available for circular constructions involving belief and believability. In particular, it would be interesting to see what would happen if there were a belief operator[10] in the language and a couple of rules for it. If a general purge of self-reference is possible, there should be a way of proving the paradoxicality of the Believer and the Modal Believer (the sentence that says of itself that it is not rationally believed and the sentence that says of itself that it is not rationally believable, respectively) in the extended system. However, things are problematic for belief. In particular, if we mimic Cook's system for knowability we would have the following two rules:

Rule B1: if $T(S_m) \vdash T(S_n)$ and $T(S_m) \vdash \neg\Diamond BT(S_n)$, then $\vdash \neg\Diamond BT(S_m)$

Rule B2: if $T(S_m) \vdash \bot$, then $\vdash \neg\Diamond BT(S_m)$

Unfortunately, these rules are much more controversial than the ones for knowability. B2, for example, implies there cannot be an agent and a time such that the agent rationally believes at that time something that logically leads to a contradiction. This strikes me as too strong.

Be that as it may, this does not imply that that there is no other way in which the purge of self-reference could be carried out. In

[9] Cook uses a sequent calculus. Here I will work with a Fitch-style natural deduction system.

[10] I am dealing with rational belief, so the intended reading of B is 'at some time someone rationally believes'.

particular, it is not difficult to see that if some reasonable principles concerning □ and K (and B) are accepted, the paradoxicality of several (all?) modal-epistemic paradoxes and their unwindings is provable without resorting to multi-modal rules. The same holds of the usual paradoxes involving just one modality, such as the Liar, the Knower and the Believer.

Let us assume that we have the following unimodal principles and rules for □ and K[11]:

(□-fact)	$\Box\Phi \rightarrow \Phi$	(K-fact)	$K\Phi \rightarrow \Phi$
(□-nec)	If $\vdash \Phi$, then $\vdash \Box\Phi$	(K-nec)	If $\vdash \Phi$, then $\vdash K\Phi$
(□-dist)	$\Box(\Phi \rightarrow \Psi) \rightarrow (\Box\Phi \rightarrow \Box\Psi)$	(K-dist)	$K(\Phi \rightarrow \Psi) \rightarrow (K\Phi \rightarrow K\Psi)$

The following is a simple proof of the paradoxicality of the Modal Knower in which only these unimodal rules are used. Proof: Assume $KT(S_1)$, where $\delta(S_1) = \neg\Diamond KT(S_1)$. By (K-fact), we can infer $T(S_1)$. Using δ-intro, we obtain $\neg\Diamond KT(S_1)$, which is equivalent to $\Box\neg KT(S_1)$. Using (□-fact) we reach a contradiction between $\neg KT(S_1)$ and $KT(S_1)$. By *reductio*, we have $\neg KT(S_1)$. Applying (□-nec) yields $\Box\neg KT(S_1)$, from which we infer $\neg\Diamond KT(S_1)$. By δ-elim we obtain $T(S_1)$. At this point we can apply (K-nec), which gives us $KT(S_1)$. And this contradicts $\neg KT(S_1)$.

Next we show that if the modal-epistemic logic for K and $\vdash$ has the unimodal rules specified above, the paradoxicality of the unwinding of the Modal Knower is provable as well. Proof:

1)	$T(S_{<1,1>})$	Assumption
2)	$\wedge\{\neg\Diamond KT(S_{<n,1>})$ for every $n > 1\}$	δ-intro
3)	$\neg\Diamond KT(S_{<2,1>})$	^-elim
4)	$\neg\Diamond KT(S_{<3,1>})$	^-elim
….	….	
ω)	$\wedge\{\neg\Diamond KT(S_{<n,1>})$ for every $n > 2\}$	^-intro
ω+1)	$T(S_{<2,1>})$	δ-elim
ω+2)	$T(S_{<1,1>}) \supset T(S_{<2,1>})$	⊃-intro
ω+3)	$K\,(\,T(S_{<1,1>}) \supset T(S_{<2,1>})\,)$	(K-nec)
ω+4)	$KT(S_{<1,1>}) \supset KT(S_{<2,1>})$	(K-dist)
ω+5)	$T(S_{<1,1>})$	Assumption
ω+6)	$\wedge\{\neg\Diamond KT(S_{<n,1>})$ for every $n > 1\}$	δ-intro

[11] For belief I assume (B-dist), (B-nec) plus the following principle that plausibly holds of an idealized notion of belief:

(B-semifact) $B\neg B\Phi \rightarrow \neg B\phi$

ω+7)	$\neg\Diamond KT(S_{<2,1>})$	$\wedge$-elim
ω+8)	$\Box\neg KT(S_{<2,1>})$	$\Box\neg =_{df} \neg\Diamond$
ω+9)	$\neg KT(S_{<2,1>})$	($\Box$-fact)
ω+10)	$T(S_{<1,1>}) \supset \neg KT(S_{<2,1>})$	$\supset$-intro
ω+11)	$KT(S_{<1,1>}) \supset T(S_{<1,1>})$	(K-fact)
ω+12)	$KT(S_{<1,1>}) \supset \neg KT(S_{<2,1>})$	Transitivity of $\supset$
ω+13)	$\neg KT(S_{<1,1>})$	$\neg$-intro
ω+14)	$\Box\neg KT(S_{<1,1>})$	($\Box$-nec)
ω+15)	$\neg\Diamond KT(S_{<1,1>})$	$\Box\neg =_{df} \neg\Diamond$
….	….	
2ω+15)	$\neg\Diamond KT(S_{<2,1>})$	$\Box\neg =_{df} \neg\Diamond$
….	….	
3ω+15)	$\neg\Diamond KT(S_{<3,1>})$	$\Box\neg =_{df} \neg\Diamond$
….	….	
ω^2)	$\wedge\{\neg\Diamond KT(S_{<n,1>})$ for every $n > 1\}$	$\wedge$-intro
ω^2+1)	$T(S_{<1,1>})$	δ-elim
ω^2+2)	$KT(S_{<1,1>})$	(K-nec)
ω^2+3)	$\Box\neg KT(S_{<1,1>}) \supset \neg KT(S_{<1,1>})$	($\Box$-fact)
ω^2+4)	$\neg\Box\neg KT(S_{<1,1>})$	modus tollens
ω^2+5)	$\Diamond KT(S_{<1,1>})$	$\Box =_{df} \neg\Diamond\neg$
ω^2+6)	$\perp$	$\perp$-intro[12]

Let me finish by saying what remains to be done. If Sorensen's idea is to be fully vindicated by means of the strategy presented above, some conditions must be set up on what counts as an adequate unwinding. A very reasonable demand is that the unwinding should have the same semantic status as the construction being unwinded. This is what Cook calls the "Covariation criterion". Informally, we are demanding that a construction C should be semantically similar to its unwinding. This requires, *at least*, that a construction C be paradoxical/non-paradoxical if and only if its unwinding is paradoxical/non-paradoxical. I believe that a lot of work has to be done in order to show that this criterion is satisfied, especially for a language like $L_{\Box KT(x)}$: we need an adequate interpretation for an *infinitary* language that has a truth *predicate* and operators for necessity, knowledge, and maybe other notions too. Also, we need to come up with a way of interpreting the language such that every construction

[12] There are analogous proofs for other modal-epistemic paradoxes (and their unwindings), such as the ◊K-version of Curry's sentence, the Knower, the Believer, the Modal Believer, etc. Obtaining an inconsistency from the ones involving belief (and their unwindings) requires (B-semifact).

is semantically similar (in a relevant way) to its unwinding. This is not to say that the project is not viable, I am just pointing out what remains to be done if the strong version of the purge is to be vindicated.

A second reasonable demand is that no sentence in the unwinding should be circular. We could call this demand the "Failure of Circularity Criterion". I have already mentioned that there are several ways to flesh this out. If circularity is to be coded as being a fixed point (of some sort), then the idea is that no sentence in the unwinding should be a fixed point (of that sort). If circularity is explained in some other way, then the criterion must be characterized differently.

One positive aspect of infinitary languages (lacking quantifiers) is that it is not hard to provide an acceptable definition of circularity for its sentences (since the difficulties are usually associated with quantified statements)[13]. Additionally, it is important to notice that I need something slightly weaker than a definition. It would be enough to provide a sufficient condition for non-circularity (or a necessary condition for circularity) and to show that the infinitary unwindings always satisfy that condition.

References

Cook, R. (forthcoming), *The Yablo Paradox: An Essay on Circularity,* Oxford, Oxford University Press.

Schlenker, P. (2007), "The Elimination of Self-Reference: Generalized Yablo Series and the Theory of Truth", *Journal of Philosophical Logic,* 36, pp. 251-307.

Sorensen, R. (1998), "Yablo's Paradox and Kindred Infinite Liars", *Mind,* 107, pp. 137-155.

[13] For instance, since $L_{\Box KT(x)}$ has no quantifiers, we could use something like Picollo's notion of m-circularity to give a full account of circularity in this language (see Picollo's contribution to this volume).