**Paradox without satisfaction**

Otávio Bueno & Mark Colyvan

1. Introduction

Consider the following denumerably infinite sequence of sentences:

\[(s_1) \text{ For all } k > 1, s_k \text{ is not true.}\]
\[(s_2) \text{ For all } k > 2, s_k \text{ is not true.}\]
\[(s_3) \text{ For all } k > 3, s_k \text{ is not true.}\]
\[\ldots\]
\[(s_n) \text{ For all } k > n, s_k \text{ is not true.}\]
\[\ldots\]

According to Stephen Yablo, the above list generates a liar-like paradox without circularity (see Yablo 1985 and 1993). After all, in contrast with the usual liar paradox, no sentence in the Yablo list refers to itself, and as opposed to well-known liar cycles,\(^1\) no sentence refers to sentences above it in the list. However, similarly to the (usual or cyclic) liar, a contradiction is derivable from the list.

But is Yablo’s paradox really non-circular? To this question, Graham Priest gave a surprising answer. In his view, despite initial appearances, the paradox is circular (Priest 1997). After all, if the formulation of the paradox doesn’t seem to involve circularity, the argument to contradiction definitely does. As Priest argued, Yablo’s paradox has ‘a fixed point … of exactly the same self-referential kind as in the liar paradox’. As a result, ‘the circularity is … manifest’ (Priest 1997: 238).\(^2\)

In this paper, we challenge Priest’s answer – in a new way. To the best of our knowledge, everyone in the debate has conceded the adequacy of Priest’s reconstruction of the argument to contradiction. We point out a limitation that hasn’t been noted. Priest’s argument requires the existence of a satisfaction relation that plays the role of a fixed point in Yablo’s paradox. However, if a contradiction can be established from the Yablo list without invoking such a relation, there’s no fixed point, and so Priest’s argument is blocked. As we show below, as opposed to Priest’s claim, the argument to contradiction doesn’t require the satisfaction relation; in fact, the argument goes through perfectly well without the latter. We conclude the paper by mentioning a consequence of this way of presenting the argument to contradiction for the significance of Yablo’s paradox.

Before proceeding, it’s important to note that so far the only argument for the circularity of the Yablo list is Priest’s. But, as hinted above, and as will become clear in a moment, Priest obtains the fixed point from the argument to contradiction, not from the construction of the list. So, as far as Priest’s argument is concerned, what we have to say here stands even if the fixed point turns out to be a feature of the list. After all, as we argue, at least the argument to contradiction doesn’t require a fixed point. Now, whether the fixed point is a feature of the list or not is hard to tell, given that, to the best of our knowledge, no argument has been presented to support this claim. Our focus here is on Priest’s challenge – the only challenge so far to the non-circularity of Yablo’s paradox.

2. Priest’s argument

First, let’s review Priest’s argument to contradiction. The sentences in the Yablo list can be formalized with a truth predicate, \(T\), in the following way: for all natural numbers \(n\), \(s_n\) is the sentence. \(\forall k > n, \neg T s_k\). Priest’s argument to contradiction goes as follows (Priest 1997: 237): For every \(n\),

\[T s_n \Rightarrow \forall k > n, \neg T s_k \quad (*\)]
\[\Rightarrow \neg T s_{n+1}. \quad (\star)\]

But,

\[T s_n \Rightarrow \forall k > n, \neg T s_k \quad (*\)]
\[\Rightarrow \forall k > n + 1, \neg T s_k \]
\[\Rightarrow T s_{n+1}. \quad (**\)]

Thus, given that ‘\(T s_n\)’ entails a contradiction, we conclude that \(\neg T s_n\). Priest then notes:

But \(n\) was arbitrary. Hence \(\forall k \neg T s_k\), by Universal Generalization. In particular, then, \(\forall k > 0, \neg T s_k\), i.e., \(s_0\), and so \(T s_0\). Contradiction (since we have already established \(\neg T s_0\)). (Priest 1997: 237)\(^3\)

Having established the contradiction, Priest invites us to focus on the lines marked ‘\((*)\)’, and asks about their justification. He claims:

\(^3\) Note the use of universal generalization and instantiation in this passage. We will return to this point later.
It is natural to suppose that this is the $T$-schema, but it is not. The $n$ involved in each step of the reduction argument is a free variable, since we apply universal generalization to it a little later; and the $T$-schema applies only to sentences, not to things with free variables in it. It is nonsense to say, for example, ‘$T'x$ is white’ iff $x$ is white. What is necessary is, of course, the generalization of the $T$-schema to formulas containing free variables. ... This involves the notion of satisfaction. For the lines marked (*) to work, they should therefore read:

\[
S(n, s') \Rightarrow \forall k > n, \neg Ts_k
\]

where $S$ is the two-place satisfaction relation between numbers and predicates, and $s'$ is the predicate $\forall k > x, \neg Ts_k$ (Priest 1997: 237)

A similar point also applies to the line marked (**). In this case, what we need is:

\[
\forall k > n + 1, \neg Ts_k \Rightarrow S(n + 1, s').
\]

And by rewriting every other line of the argument accordingly, replacing truth by satisfaction, the final contradiction $\neg \exists k > 0, \neg S(k, s')$ and its negation – is obtained.

It’s now clear, Priest concludes, that the paradox concerns a predicate, $s'$, of the form $\forall k > x, \neg S(k, s')$; and the fact that $s' = \forall k > x, \neg S(k, s')$ shows that we have a fixed point, $s'$, here, of exactly the same self-referential kind as in the liar paradox.

In a nutshell, $s'$ is the predicate ‘no natural number greater than $x$ satisfies this predicate’. The circularity is now manifest. (Priest 1997: 238)

Even if we grant that the circularity is now manifest, the question still arises: is the circularity inherent in Yablo’s paradox, or is it simply an artefact of the particular version of the argument to contradiction used by Priest? We think that the latter is the case.

3. The argument reformulated

We now show how to derive a contradiction from the Yablo list, without applying the $T$-schema to any open formula. And so there is no need to invoke a satisfaction relation to run Yablo’s paradox. As will become clear, our argument is parasitic on Priest’s argument, but with a crucial new twist.

Consider ‘$s_1$’ in the Yablo list. Suppose ‘$s_1$’ is true (which, as above, we denote by ‘$T_s$’).

\[
T_{s_1} \Rightarrow \forall k > 1, \neg Ts_k
\]

\[
\Rightarrow \neg Ts_1
\]

But,

\[
T_{s_1} \Rightarrow \forall k > 1, \neg Ts_k
\]

\[
\Rightarrow \forall k > 2, \neg Ts_k
\]

\[
\Rightarrow \neg Ts_2
\]

So, given that ‘$T_{s_1}$’ entails a contradiction, $\neg Ts_1$. This means that there is at least one true sentence in the Yablo list. Let the first such sentence be ‘$s_1$’. (Note that ‘$1$’ is not a variable, but an unknown, particular natural number.) Now consider ‘$s_1$’.

\[
T_{s_1} \Rightarrow \forall k > 1, \neg Ts_k
\]

\[
\Rightarrow \neg Ts_1
\]

But,

\[
T_{s_1} \Rightarrow \forall k > 1, \neg Ts_k
\]

\[
\Rightarrow \forall k > 1 + 1, \neg Ts_1
\]

\[
\Rightarrow Ts_{s_1}
\]

Thus, a contradiction can be derived from the truth or untruth of a particular sentence, ‘$s_1$’, in the Yablo list.

Three remarks are in order here:

(a) The sentence we used to establish the contradiction is, of course, arbitrary – we could have started with any sentence in the list. But – and this is crucial – the arbitrariness of the sentence we started with is never used to derive the contradiction, and so it’s not necessary for the argument to go through. Even if ‘$s_1$’ were the only paradoxical sentence in the Yablo list, this would be sufficient to conclude that Yablo’s paradox (i) is a paradox, and (ii) is not circular – or, at least, it’s not circular in the sense at issue here (i.e. it doesn’t require a fixed point of a self-referential kind).

(b) Priest’s argument to contradiction is unnecessarily strong. The argument actually establishes that every sentence in the Yablo list is paradoxical. As we saw, to get this conclusion, Priest needs to use universal generalization at a crucial point in the argument, just to instantiate the result in the following line (see our first full quotation from Priest’s paper in §2). By focusing on a particular sentence in the Yablo list, and then only establishing that there is at least one paradoxical sentence in the list, our argument bypasses this move altogether. After all, we don’t need to prove that Yablo’s paradox is massively paradoxical to establish that it is a paradox!

(c) Finally, in our argument, we have not used the $T$-schema on anything other than sentences in order to derive the contradiction. In particular, we have not illegitimately used the $T$-schema on any open formulae. As we saw, the contradiction can be derived from the Yablo list without invoking a satisfaction relation, and so the fixed point is only an artefact of the argu-

\footnote{We have made minor changes to Priest’s notation, but nothing hangs on this.}
Tooley on backward causation

Paul Noordhof

Michael Tooley has argued that, if backward causation (of a certain kind) is possible, then a Stalnaker-Lewis account of the truth conditions of counterfactuals cannot be sound. I shall argue that he is wrong.¹ According to David Lewis,

A counterfactual 'If it were that A, then it would be that C' is non-vacuously true if and only if some (accessible) world where both A and C are true is more similar to our actual world, over-all, than is any world where A is true but C is false. (Lewis 1979: 41)

Lewis’s criteria for assessing the similarity between possible worlds are as follows.

(A) It is of the first importance to avoid big, widespread, diverse violations of law.
(B) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
(C) It is of the third importance to avoid even small, localized, simple violations of law.
(D) It is of little or no importance to secure approximate similarity of particular fact, even in matters which concern us greatly. (Lewis 1979: 47–48)

The basic idea is that we are to consider all those worlds in which A is true. The counterfactual will be true if the worlds in which C is also true are more similar according to the criteria laid out in (A) to (D) than any world in which C is not true. The crucial point is that which close worlds we consider is fixed by, in the first instance, the envisaged truth of the antecedent.

Tooley invites us to imagine a world in which the following holds.

Law 1: For any location x, and time, t, if location x has both property P and property Q at time t, then that state of affairs causes a related location x + Δx to have property P, and to lack property Q, at the later time t + Δt.

¹ Initially, Tooley suggests that he shall demonstrate that backward causation is incompatible with the Stalnaker-Lewis style account of the truth conditions of counterfactuals (Tooley 2002: 191). By the end, it becomes clear that he only thinks that the Stalnaker-Lewis style account of the truth conditions of counterfactuals is incompatible with backward causation in worlds whose laws rule out causal loops.