Yablo's paradox

Graham Priest

1. *Introduction*

Stephen Yablo has given an ingenious liar-style paradox that, he claims, avoids self-reference, even of an indirect kind, one that is, in fact, 'not in *any* way circular' (Yablo 1993, his italics). He infers that such circularity is not necessary for this kind of paradox. Some others have agreed.[1] The point of this note is to demonstrate that self-referential circularity *is* involved in Yablo's paradox. I shall also show that Yablo's paradox has exactly the same structure as all the familiar paradoxes of set theory and semantics.

To put the discussion into context, think, first, of the standard Liar paradox, 'This sentence is not true'. Writing T as the truth predicate, then the Liar sentence is one, t, such that t = '$\neg Tt$'. The fact that 't' occurs on both sides of the equation makes it a fixed point of a certain kind, and, in this context, codes the self-reference.

In this particular case, the existence of the fixed point is obvious, due to the use of the demonstrative 'This sentence', but in general one often has to work quite hard to show the existence of fixed points. For example, if we take naming to be implemented by gödelization, we have to prove the existence of a certain number. Specifically, we show, by a diagonalization argument, that if $\alpha(x)$ is any formula of one free variable, x, there is a number, n, such that n is the code of the formula $\alpha(\underline{n})$ – where $\underline{n}$ is the numeral of n – or, at least, of one logically equivalent to it. n is the fixed point.[2]

2. *Yablo's paradox*

Bearing this in mind, let us now turn to Yablo's paradox itself. Yablo sets this up with the following words:[3]

[1] E.g., Tennant 1995, Sorensen 199+. Versions of Yablo's argument are also to be found in Yablo 1985: 340, and Visser 1989: sect. 3.2.2.

[2] See, for example, Boolos and Jeffrey 1974: ch. 15. In mathematics, a fixed point for a function, f, is some x such that $x = f(x)$. Strictly speaking, the t and n just mentioned are not fixed points in this sense, since they occur in opaque contexts as arguments. What n is a fixed point of, is the function f_α which, for any input, m, gives the sentence which is $\alpha(x)$ with the canonical name for m substituted for x. $n = f_\alpha(n)$. Similarly for t.

[3] Yablo uses an upper case S. For reasons that will become clear, I will use lower case s. Yablo also counts from 1 rather than 0, but nothing hangs on this.

Analysis 57.4, October 1997, pp. 236–242. © Graham Priest

Imagine an infinite sequence of sentences s_0, s_1, s_2, ..., each to the effect that every subsequent sentence is untrue:

(s_0) for all $k > 0$, s_k is untrue,
(s_1) for all $k > 1$, s_k is untrue,
(s_2) for all $k > 2$, s_k is untrue, ...

Formalizing the sentences with a truth predicate, T, we have that for all natural numbers, n, s_n is the sentence $\forall k > n, \neg Ts_k$. Note that each sentence refers to (quantifies over) only sentences later in the sequence. No sentence, therefore, refers to itself, even in an indirect, loop-like, fashion. There seems to be no circularity.

Given this set-up, the argument to contradiction goes as follows. For any n:

$$\begin{aligned} Ts_n &\Rightarrow \forall k > n, \neg Ts_k && (*) \\ &\Rightarrow \neg Ts_{n+1} \end{aligned}$$

But:

$$\begin{aligned} Ts_n &\Rightarrow \forall k > n, \neg Ts_k && (*) \\ &\Rightarrow \forall k > n{+}1, \neg Ts_k \\ &\Rightarrow Ts_{n+1} && (**) \end{aligned}$$

Hence, Ts_n entails a contradiction, so $\neg Ts_n$. But n was arbitrary. Hence $\forall k \neg Ts_k$, by Universal Generalization. In particular, then, $\forall k > 0, \neg Ts_k$, i.e., s_0, and so Ts_0. Contradiction (since we have already established $\neg Ts_0$).

Now, focus on the lines marked (*). What is their justification? It is natural to suppose that this is the T-schema, but it is not. The n involved in each step of the reductio 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 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. (For the purposes of this paper, I will call such things 'predicates'.) This involves the notion of satisfaction. For the lines marked (*) to work, they should therefore read:

$$S(n, \dot{s}) \Rightarrow \forall k > n, \neg Ts_k$$

where S is the two-place satisfaction relation between numbers and predicates, and $\dot{s}$ is the predicate $\forall k > x, \neg Ts_k$.

A similar comment applies to the line marked (**). For this to work, it should be:

$$\forall k > n{+}1, \neg Ts_k \Rightarrow S(n{+}1, \dot{s})$$

But then every other line of the argument needs to be rewritten to make it work, truth being replaced by satisfaction. In particular, $\dot{s}$ has to be taken as the predicate $\forall k > x, \neg S(k, \dot{s})$. Rewriting in this way, the argument goes through straightforwardly, as may be checked. The final contradiction is $\forall k > 0, \neg S(k, \dot{s})$ and its negation.

None of this is profound, but it focuses attention on the fact that the paradox concerns a predicate, $\dot{s}$, of the form $\forall k > x, \neg S(k, \dot{s})$; and the fact that $\dot{s}$ = '$\forall k > x, \neg S(k, \dot{s})$' shows that we have a fixed point, $\dot{s}$, here, of exactly the same self-referential kind as in the liar paradox. In a nutshell, $\dot{s}$ is the predicate 'no number greater than x satisfies this predicate'. The circularity is now manifest.

The existence of the fixed point predicate is guaranteed, as in the liar case, by the demonstrative. If naming were implemented with arithmetic, to establish the existence of the fixed point, we would need a standard generalization of the diagonal argument, to the effect that if $\alpha(x, y_1, \ldots, y_m)$ is any formula with the free variables displayed, there is a number, n, which is the code number of $\alpha(\underline{n}, y_1, \ldots, y_m)$ (or of a formula logically equivalent to it).[4]

This answers a question that should have been obvious as soon as one reads Yablo's description of the situation. He asks us to imagine a certain sequence. How can one be sure that there is such a sequence? (We can imagine all sorts of things that do not exist.) As he presents things, the answer is not at all obvious. In fact, we can be sure that it exists because it can be defined in terms of $\dot{s}$: the n-th member of the sequence is exactly the predicate $\dot{s}$ with 'x' replaced by '$\underline{n}$'.[5]

3. *Infinitary Reasoning*

As Yablo's version of the argument stands, the demonstration that it involves self-reference is definitive. Is there any hope of reformulating it in such a way as not to involve self-reference?

One might suggest the following. We leave the deduction as first laid out, but construe the n in the reductio part of the argument as schematic, standing for any natural number. This gives us an infinity of proofs, one of $\neg Ts_n$, for each n. We may then obtain the conclusion $\forall n \neg Ts_n$ by an application

[4] Using this, one can turn Yablo's argument into a proof of Gödel's first incompleteness theorem by considering the predicate $\forall k > y, \neg\exists y \pi(y, x, k)$, where $\pi(y, x, k)$ is the primitive recursive predicate: y is (the code of) a proof of the predicate (with code) x when the numeral for k is substituted for its free variable. We construct the fixed point predicate, and then show that it can be neither proven nor refuted of 0, by Yabloesque reasoning.

[5] The problem of showing existence is even clearer in a set-theoretic analogue of Yablo's paradox given by Goldstein 1995. This concerns a sequence of sets, C_n, for each natural number, n, satisfying the condition:

$$\forall x(x \in C_n \leftrightarrow \forall k > n, x \notin C_k)$$

Given the non-well-foundedness of the situation, it is not at all clear that there is such a sequence, even in naive set theory with an unrestricted comprehension principle.

of the ω-rule:

$$\frac{\alpha(0),\ \alpha(1),\ \ldots}{\forall x\alpha(x)}$$

The rest of the argument is as before. Construing the argument in this way, we do not have to talk of satisfaction. There is therefore no predicate involved, and a fortiori no fixed-point predicate. We therefore have a paradox without circularity.[6]

Such a suggestion would be disingenuous, though. As a matter of fact, we did not apply the ω-rule, and could not have. The reason we know that $\neg Ts_n$ is provable for all n is that we have a uniform proof, i.e., a proof for *variable n*. Moreover, no finite reasoner ever really applies the ω-rule. The only way that they can know that there is a proof of each $\alpha(i)$ is because they have a uniform method of constructing such proofs. And it is this finite information that grounds the conclusion $\forall x\alpha(x)$.

Still, it might be suggested, at least for an infinite being, God, say, who really can apply the ω-rule, there is a paradox here that does not involve circularity. Even this is false, however. I chose to demonstrate that Yablo's paradox involves circularity by analysing a detail of the argument involved, since this brings out the circularity most clearly. However, the circularity has nothing to do with the *argument* as such; it arises in the *structure* of the situation. And this is equally true, though perhaps less obvious, of the original formulation. The paradox concerns a sequence of sentences, s_n, or $s(n)$, to remind the reader that the subscript notation is just a notational variant of a function applied to its argument. The function s is defined by specifying each of its values, but each of these is defined with reference to s. (As a glance at Yablo's original formulation suffices to demonstrate.) It is now the function s that is a fixed point. s is the function which, applied to any number, gives the claim that all claims obtained by applying *s itself* to subsequent numbers are not true. Again, the circularity is patent. Note the role that the infinite regress in playing here. If the regress grounded out in some claim not concerning the sequence, then s could be defined recursively, and it would not require a circular construction to define it. But the sequence is infinite; and it does.[7]

Given infinitary fantasies, the circularity can be further masked. Consider, for example, the following version of the paradox (due, essentially to Sorensen 199+). At the gates of Heaven an infinite queue of people

[6] The argument is formulated in an infinitary fashion in Hardy 1995.

[7] Nor would it help to employ infinitary connectives, as Thomas Forster formulates the paradox in 199+. If we do, we may avoid talk of truth and satisfaction altogether, since $s(n)$ can just be taken to be $\bigwedge_{m>n} \neg s(m)$. But s is still getting in on both sides of the act.

is tailed back. Each of them is thinking one thought – and it's not 'Will there be room for me here?'. It is: the thought[8] that each person behind me is thinking is not true. Now, God, it would appear, can reason about every person, and deduce a contradiction as before. At first glance, there would appear to be no circularity here.

But there is. This is most obvious if one individuates thoughts in such a way that all the people are thinking the same thought, *t*. If this is the case, then the thought that they are thinking is just equivalent to the thought that *t* is not true. The circularity is obvious. In fact, this is just a variant of the liar paradox. Even if thoughts are individuated in such a way that the people may be thinking different thoughts,[9] circularity is still present. Let *x*'s thought be $t(x)$. What is $t(x)$? Simply: $\forall y$(if *y* is behind *x* then $t(y)$ is not true). *t* is that function whose value at *x* is the claim that, for every *y*, behind *x*, the value of *t itself* applied to *y*, is not true.[10] In the language of the lambda calculus: $t = \lambda x \forall y$(if *y* is behind *x* then $t(y)$ is not true). The circularity implicit in the structure in clear.

The implicit nature of the circularity in this version of the paradox is a distinctive feature. But this is not a novelty. Consider Prior's 1961 example of the man thinking that nothing now being thought in room 7 is true – when, unbeknown to him, he himself is in room 7. There is nothing explicitly circular in the man's thought. Rather, the circularity arises implicitly due to the geographical configuration. Similarly in Sorensen's version of Yablo's paradox. In fact, Sorenson's version of the paradox is, essentially, to Prior's what Yablo's is to the Liar.

4. *The Inclosure Schema*

The situation involved in Yablo's paradox, however formulated, is intrinsically circular, in exactly the same way that those involved in more familiar paradoxes of the family are. It is not, therefore, surprising that the paradox has exactly the same structural characterization as all these other paradoxes, as I will now show.

An *inclosure* is a triple $\langle \delta, \Omega, \theta \rangle$, where Ω is a set of objects, θ is a property defined on subsets of Ω, such that $\theta(\Omega)$, and δ is a partial function from subsets of Ω to Ω, defined on the sets of which θ is true, and such that if $X \subseteq \Omega$:

[8] Or every thought; the argument works just as well.

[9] Or we ensure that they are, by getting each person to add a conjunct of the form $n = n$ to their thought, where they are *n*-th in line.

[10] Or if we take the paradox in the 'every' form, $t(x)$ is the set of *x*'s thoughts, and $t(x)$ is the set whose only member is $\forall y \forall z$(if *y* is behind *x* and $z \in t(y)$ then *z* is not true).

$$\delta(X) \notin X \qquad \text{(Transcendence)}$$
$$\delta(X) \in \Omega \qquad \text{(Closure)}$$

Loosely speaking, δ is a function that produces an object which 'diagonalizes' out of X whilst remaining in Ω (for subsets of Ω in the family characterized by θ). Contradiction arises if we apply δ to the limit structure, Ω, itself. For we then get $\delta(\Omega) \notin \Omega$ but $\delta(\Omega) \in \Omega$. In Priest 1995 it is shown that every standard paradox of self-reference (whether semantic or set-theoretic) has the structure of an inclosure. I will not repeat the details here.

To see that Yablo's paradox has the same structure, let $\Omega = \{\langle n, p\rangle; S(n, p)\}$, where S is, as before, the satisfaction relation between numbers and one-place arithmetic predicates. Let θ be the property of being definable, i.e., $\theta(X)$ iff there is a name that refers to X; and define δ on definable subsets of Ω as follows. $\delta(X)=\langle 0, r_X\rangle$, where r_X is the predicate:

$$\dot{s} \wedge \forall k > 0, \langle k, \dot{s}\rangle \notin \underline{X}$$

$\underline{X}$ is a name for X and $\dot{s}$, recall, is the predicate $\forall k > X, \neg S(k, \dot{s})$.[11]

It is clear that Ω is definable (I have just defined it); hence, $\theta(\Omega)$. To show that the structure is an inclosure, it remains to check Transcendence and Closure. The arguments are, unsurprisingly, simply the arguments of Yablo's paradox. Suppose that X is definable, and that $X \subseteq \Omega$. For Transcendence:

$$\begin{aligned}
\delta(X) \in X &\Rightarrow \langle 0, r_X\rangle \in X \\
&\Rightarrow S(0, r_X) \\
&\Rightarrow \forall k > 0, \neg S(k, \dot{s}) \wedge \forall k > 0, \langle k, \dot{s}\rangle \notin \underline{X} \\
&\Rightarrow \forall k > 0, \neg S(k, \dot{s}) \\
&\Rightarrow \neg S(1, \dot{s}) \wedge \forall k > 1, \neg S(k, \dot{s}) \\
&\Rightarrow \neg S(1, \dot{s}) \wedge S(1, \dot{s})
\end{aligned}$$

Hence, $\delta(X) \notin X$.

For Closure, we need to show that $\delta(X) \in \Omega$, i.e., $S(0, r_X)$, i.e., $\forall k > 0$, $\neg S(k, \dot{s}) \wedge \forall k > 0, \langle k, \dot{s}\rangle \notin \underline{X}$. Since $\langle k, \dot{s}\rangle \in \underline{X}$ entails $S(k, \dot{s})$, $\neg S(k, \dot{s})$ entails $\langle k, \dot{s}\rangle \notin \underline{X}$, and the first conjunct entails the second. It therefore suffices to prove the first conjunct.

$$\begin{aligned}
\neg\forall k > 0, \neg S(k, \dot{s}) &\Rightarrow \exists k > 0, S(k, \dot{s}) \qquad \text{Let this } k \text{ be } n. \\
&\Rightarrow S(n, \dot{s}) \\
&\Rightarrow \forall k > n, \neg S(k, \dot{s}) \\
&\Rightarrow \neg S(n+1, \dot{s}) \wedge \forall k > n+1, \neg S(k, \dot{s}) \\
&\Rightarrow \neg S(n+1, \dot{s})) \wedge S(n+1, \dot{s})
\end{aligned}$$

[11] The second conjunct is, in fact, logically unnecessary. But it ensures that $\delta(X)$ depends genuinely on X.

Hence, by reductio, $\forall k > 0, \neg S(k, \dot{s})$.

We see that the structure is an inclosure. The contradictory sentence concerning the limit is: $\delta(\Omega) \in \Omega$, i.e., $\langle 0, r_\Omega \rangle \in \Omega$, i.e., $S(0, r_\Omega)$, i.e., $\forall k > 0$, $\neg S(k, \dot{s}) \wedge \forall k > 0, \langle k, \dot{s} \rangle \in \underline{\Omega}$. But the second conjunct is equivalent to the first. Hence the paradoxical sentence is $\forall k > 0, \neg S(k, \dot{s})$, which is that of Yablo's paradox.

5. *Conclusion*

As we see, then, Yablo's paradox does involve circularity of a self-referential kind. However one formulates it, it has the characteristic fixed-point structure. Moreover, the paradox is an inclosure contradiction, as are all the other paradoxes of its kind. Yablo's paradox is therefore just another of the same family. This is not to belittle it. The family is a rich and varied one. It contains paradoxes as different in detail as the Liar paradox, Cantor's paradox and Berry's paradox. Yablo's paradox demonstrates a new variety of this richness.[12]

University of Queensland
Brisbane, Australia 4072
ggp@lingua.cltr.uq.oz.au

References

Boolos, G. and R. Jeffrey. 1974. *Computability and Logic*. Cambridge: Cambridge University Press.

Forster, T. 199+. The significance of Yablo's paradox without self-reference. Typescript.

Goldstein, L. 1994. A Yabloesque paradox in set theory. *Analysis* 54: 223–27.

Hardy, J. 1995. Is Yablo's paradox liar-like? *Analysis* 55: 197–98.

Priest, G. 1995. *Beyond the Limits of Thought*. Cambridge: Cambridge University Press.

Prior, A. 1961. On a family of paradoxes. *Notre Dame Journal of Formal Logic* 2: 16–32.

Sorensen, R. 199+. Extroverted epistemic paradox. Typescript.

Tennant, N. 1995. On paradox without self-reference. *Analysis* 55: 199–207.

Visser, A. 1989. Semantics and the liar paradox. In *Handbook of Philosophical Logic*, Vol. IV, eds. D. Gabbay and F. Guenthner, 617–706. Dordrecht: Kluwer Academic Publishers.

Yablo, S. 1985. Truth and reflection. *Journal of Philosophical Logic* 14: 297–349.

Yablo, S. 1993. Paradox without self-reference. *Analysis* 53: 251–52.

[12] I would like to thank Len Goddard and, particularly, Roy Sorensen for a number of interesting discussions of Yablo's paradox. A draft of this paper was read to the Philosophy Department at Monash University. I am grateful to those present for helpful comments, and particularly to Alan Hazen, Richard Holton and Lloyd Humberstone.