

Bueno and Colyvan on Yablo's paradox

JEFFREY KETLAND

1. Introduction

Yablo's paradox (Yablo 1993) involves a denumerable sequence Y_i of sentences, with the following truth conditions:

Y_0 : For all $n > 0$, Y_n is not true.

Y_1 : For all $n > 1$, Y_n is not true.

etc.

Unlike the paradoxicality of the strengthened Liar sentence, the paradoxicality of this infinite list of 'Yablo sentences' is not at all straightforward. In Ketland (forthcoming), it is shown that the list of Yablo sentences is *not* formally inconsistent. More precisely, the list of 'Yablo biconditionals' (all instances of ' $Y_n \leftrightarrow \forall m > n, Y_m$ is not true') is *not* inconsistent with the relevant *local* disquotation principle (all instances of ' Y_n is true $\leftrightarrow Y_n$ '). Rather, the combination is ω -inconsistent. Furthermore, with an appropriate definition of the extension of 'true', it is possible to satisfy this combination on any *non-standard* model of arithmetic.¹ So, in a sense, Yablo's paradox is not a genuine paradox. Rather, it is an ' ω -paradox', an infinite set of sentences which is unsatisfiable on the standard model of arithmetic.

2. Formalization

A natural formalization of Yablo's paradox was given by Priest 1997. This analysis is continued in Ketland (forthcoming), where six central statements and/or schemes are isolated:

(A) The Uniform Homogeneous Yablo Principle

For all n , Y_n is true if and only if, for all $m > n$, Y_m is not true.

This uniform statement is 'homogeneous' in the sense that it has the truth predicate on both sides. It is easy to show, assuming the usual properties of the linear order $>$, that it is inconsistent. The *logical form* of the Uniform Homogeneous Yablo Principle is,

(B) Uniform Homogeneous Yablo Scheme

(B₁) $\forall x[\varphi(x) \leftrightarrow \forall y(R(y, x) \rightarrow \sim \varphi(y))]$.

¹ Stephen Yablo pointed out to me (private communication) that there is another way to see this, using a compactness argument. Each finite subset of Yablo biconditionals is satisfiable. By the compactness theorem, the whole set is satisfiable.

Add the auxiliary axioms:

$$(B_2) \quad \forall x \exists y R(y, x),$$

$$(B_3) \quad \forall x \forall y \forall z (R(x, y) \ \& \ R(y, z) \rightarrow R(x, z)).$$

It is easy to show that the set $\{B_1, B_2, B_3\}$ is inconsistent. The two latter axioms, governing the relation symbol R , intuitively express the order properties required for the inconsistency of (A).

(C) The Uniform Fixed-Point Yablo Principle

$$\forall x (Y(x) \leftrightarrow \forall y > x \sim T(\ulcorner Y(\text{dot}(y)) \urcorner)).^2$$

Using a diagonalization argument, this uniform fixed-point statement is provable in PA , in the language L_T comprising the language L of arithmetic extended with a new predicate symbol T . The L_T -formula $Y(x)$ may be called the ‘Yablo predicate’. The Yablo sentences Y_n are then $Y(\underline{n})$, where \underline{n} is any numeral. That is, the Yablo sentences are the sentences $Y(\underline{0})$, $Y(\underline{1})$, etc.

The subformula $\forall y > x \sim T(\ulcorner Y(\text{dot}(y)) \urcorner)$ is equivalent (by definitions) to the formula $\forall y > x \sim \text{Sat}(\ulcorner Y(x) \urcorner, y)$, which means ‘no number greater than x satisfies $Y(x)$ ’. The Yablo predicate $Y(x)$ is therefore a uniform fixed point of the 2-place predicate $\forall y > x \sim \text{Sat}(z, y)$, which means ‘no number greater than x satisfies z ’. All of this is explained in Priest’s 1997 article.

(D) The Local Arithmetic Disquotation Scheme

$$T(\ulcorner \varphi \urcorner) \leftrightarrow \varphi, \text{ with } \varphi \in \text{Sent}(L).$$

The *local arithmetic* disquotation scheme implements the equivalence of any *arithmetic* sentence φ and the sentence $T(\ulcorner \varphi \urcorner)$, which intuitively says that φ is true. It is well known that the local arithmetic disquotation scheme is consistent (indeed, it is conservative over reasonable arithmetic theories: see Ketland 1999).

(E) The Local Yablo Disquotation Principle

$$T(\ulcorner Y_n \urcorner) \leftrightarrow Y_n, \text{ with } n \in \omega.$$

This *local* disquotation principle implements the equivalence of any *specific* Yablo sentence Y_n and the sentence $T(\ulcorner Y_n \urcorner)$.

² This uses Feferman’s dot notation, designed to permit quantification into formulae containing quotation terms. More exactly, the term $\ulcorner \varphi(\text{dot}(x)) \urcorner$ is a function term-containing the variable x free. Intuitively it means ‘the result of substituting the numeral of the number x for all free variables in φ ’. This implies that, for example, $\ulcorner \varphi(\text{dot}(\underline{n})) \urcorner$ denotes the code of $\varphi(\underline{n})$, for each number n . The function term $\ulcorner \varphi(\text{dot}(x)) \urcorner$ is in fact definable in arithmetic, using function terms for substitution and the naming function (the function which takes a number to its canonical numeral). Note that the equation $\ulcorner \varphi(\text{dot}(\underline{n})) \urcorner = \ulcorner \varphi(\underline{n}) \urcorner$ is provable in arithmetic. So, for example, $T(\ulcorner Y(\text{dot}(\underline{n})) \urcorner) \leftrightarrow T(\ulcorner Y(\underline{n}) \urcorner)$ is provable.

(F) The Uniform Yablo Disquotation Principle

$$\forall x(T(\ulcorner Y(\text{dot}(x)) \urcorner) \leftrightarrow Y(x)).$$

This *uniform* disquotation principle is more powerful than the local disquotation principle (E). Here, x is a bound variable.

The main properties of the above statements/schemes are as follows:

- (1) The scheme (B) is inconsistent (with the axioms governing the order).
- (2) The set of numerical instances of (B) is *consistent*, albeit ω -inconsistent. Indeed, the set of numeral instances of (B) has non-standard models.
- (3) (A) implies (B). So (A) is inconsistent.
- (4) The combination (C) + (D) + (E) is *consistent*, albeit ω -inconsistent. Indeed, the combination (C) + (D) + (E) has non-standard models.
- (5) (C) + (F) implies (A). So (C) + (F) is inconsistent.

My own view, and, presumably, Priest's, is that the statements (C) and (F) provide the correct representation of Yablo's paradox, insofar as we wish to identify an underlying formal inconsistency generated by semantic principles.

3. *Bueno and Colyvan's objections*

Bueno and Colyvan 2003 have recently argued *against* the above analysis of the Yablo paradox. Replying to Priest, they write,

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. (Bueno and Colyvan 2003: 153)

In further detail, Bueno and Colyvan make three main claims:

- (i) Priest's formalization of the paradox introduces the notion of *satisfaction*. But, allegedly, this is unnecessary and 'there is no need to invoke a satisfaction relation to run Yablo's paradox'. (Bueno and Colyvan 2003: 154)
- (ii) Priest's construction of the list of Yablo sentences uses a *fixed-point construction*. But, allegedly, this is an 'artefact of the particular version of the argument to contradiction used by Priest'. (Bueno and Colyvan 2003: 154)

(iii) Allegedly, the derivation of the inconsistency from the Yablo list does not require the *uniform* disquotational T-scheme (satisfaction of open sentences). Rather, the derivation merely requires the *local* disquotation scheme.

Here we argue that each of these claims is incorrect.

In connection with (i), note first that the notions of truth and satisfaction are *interdefinable*, so long as we have a system of naming for objects. Here, we are naming numbers and formulae, and surely one's scruples about nameability shouldn't extend to such a level. We can define satisfaction in term of truth, naming and substitution as follows:

A number x *satisfies* a 1-place predicate y if and only if the result of substituting the *numeral* for x for any free variable in y is *true*.³

So, we can express the Yablo formula $Y(x)$ as follows:

For any number $y > x$, the result of substituting the numeral for y for all free variables in $Y(x)$ is not true.

And this can be simplified,

For any number $y > x$, $Y(\text{num}(y))$ is not true.

So, Priest's point about the use of satisfaction is not touched by Bueno and Colyvan's line of reasoning.

Furthermore, the Yablo formula $Y(x)$ is *obviously* a fixed-point of a certain predicate, namely, the 2-place predicate 'for any $y > x$, $z(\text{num}(y))$ is not true'. Whether one expresses this using truth or satisfaction is entirely immaterial, since they are interdefinable.

Objection (ii) appears in a footnote:

According to Graham Priest (private communication), the crucial issue in his paper is how one knows that the Yablo list exists. In addition, as he argued, to establish the list's existence, a fixed-point construction is required. However, *it's not clear to us that the argument Priest gave establishes how we know that the Yablo list exists*. Priest's argument seems to presuppose the existence of the list, in order to establish that to derive a contradiction from the latter, a fixed-point construction is required. (Bueno and Colyvan 2003: 156, italics added.)

This seems to me quite mistaken. As noted above, if one extends the language L of arithmetic by adding a 1-place predicate symbol T to form a language L_T , the uniform diagonalization theorem shows that there is an

³ More complex definitions are available for 2, 3, etc., place predicates. One also needs some way of representing finite sequences. To see this worked out in detail, see Kaye 1991.

L_T -formula $Y(x)$ such that the Uniform Fixed-Point Yablo Principle (C) is provable in PA . More precisely, we have,

$$PA \vdash \forall x [Y(x) \leftrightarrow \forall y > x \sim T(\ulcorner Y(\text{dot}(y)) \urcorner)].^4$$

The Yablo sentence Y_n is then identified with $Y(\underline{n})$, for each numeral \underline{n} . To stress, it is a *theorem of mathematical logic that the Yablo list exists*. This is a direct and well-understood construction. Priest does *not* 'presuppose the existence of the list, in order to establish that to derive a contradiction from the latter, a fixed-point construction is required'. If anything, the reverse. On Priest's analysis of the paradox, one *first* directly constructs the Yablo list, using standard diagonalization techniques, *and then* proves the inconsistency of the Uniform Fixed-Point Yablo Principle (C) with the Uniform Yablo Disquotation Principle (F).

In contrast, Bueno and Colyvan merely *stipulate* 'that s_n is the sentence $\forall k > n \sim Ts_k$ '. But (although their notation obscures this fact) this is precisely *what a fixed-point is*. If we have an mapping $f: X \rightarrow X$ and we *stipulate* that, for some $r \in X$, $r = f(r)$, then this r is a fixed point of the function f . Of course, such fixed point claims require proof, not mere stipulation.

The third claim (iii) is the most interesting and deserves a more detailed examination.

4. Bueno and Colyvan's 'derivation'

We now show how to derive a contradiction from the Yablo list, *without applying the T-schema to any open formula*. (Bueno and Colyvan 2003: 154, emphasis added)

If I understand Bueno and Colyvan correctly, this amounts to the claim that the *list* of Yablo biconditionals is deductively inconsistent with the *Local* Yablo Disquotation Principle, i.e. (E) above. Commenting on their own proof,

... 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. (Bueno and Colyvan 2003: 155)

⁴ Given a formula $\phi(x)$ with just x free, the Diagonalization Lemma shows how to construct a closed formula G such that $G \leftrightarrow \phi(\ulcorner G \urcorner)$ is provable in PA . See, e.g., Boolos, Burgess and Jeffrey 2002: 222. To formalize Yablo's paradox, we use the uniform generalization of the Diagonalization Lemma. If $\phi(z, x)$ is a formula with just z and x free, then the Uniform Diagonalization Lemma shows how to construct a formula $G(x)$ with just x free such that $\forall x (G(x) \leftrightarrow \phi(\ulcorner G(x) \urcorner, x))$ is provable in PA (actually, in Q). In the Yablo case, $\phi(z, x)$ is the formula $\forall y > x \sim Sat(z, y)$. Then $Y(x)$ is constructed such that $\forall x (Y(x) \leftrightarrow \forall y > x \sim Sat(\ulcorner Y(x) \urcorner, y))$ is provable in PA .

This cannot be right, for it can be shown that the formalized theory which proves the Yablo biconditionals and the *Local* Yablo Disquotation Principle is consistent, and has non-standard models (see Ketland forthcoming). In fact, the theory in question is a conservative extension of Peano arithmetic.

In more detail, Bueno and Colyvan's reasoning first shows that the set of Yablo biconditionals plus the *local* Yablo disquotation principle implies $\sim T(\ulcorner Y_1 \urcorner)$. In effect,

$$\{T(\ulcorner Y_1 \urcorner) \leftrightarrow Y_1, Y_1 \leftrightarrow \forall y > \underline{1} \sim T(\ulcorner Y(\text{dot}(y)) \urcorner)\} \vdash_{PA} \sim T(\ulcorner Y_1 \urcorner).$$

This is fine and similar reasoning shows that the list implies $\sim T(\ulcorner Y_n \urcorner)$ for each $n \in \omega$. That is,

$$\{T(\ulcorner Y_m \urcorner) \leftrightarrow \forall y > \underline{m} \sim T(\ulcorner Y(\text{dot}(y)) \urcorner): m \in \omega\} \vdash_{PA} \sim T(\ulcorner Y_n \urcorner), \text{ for each } n \in \omega.$$

The set of Yablo biconditionals is, as noted above, provable in Peano arithmetic. Let PA_Y be the extension of PA formed by adding the local arithmetic disquotation scheme and the local Yablo disquotation principle. The reasoning so far shows that,

$$PA_Y \vdash \sim T(\ulcorner Y(\underline{n}) \urcorner), \text{ for each } n \in \omega.$$

From this, it is then easy to deduce $\exists x > \underline{n} T(\ulcorner Y(\text{dot}(x)) \urcorner)$, for each $n \in \omega$. So,

$$PA_Y \vdash \exists x T(\ulcorner Y(\text{dot}(x)) \urcorner).$$

This implies that PA_Y is ω -inconsistent, since PA_Y proves a formula of the form $\sim\varphi(\underline{n})$, for each $n \in \omega$, but also proves $\exists x\varphi(x)$.⁵ But one *cannot* infer from this that an object which witnesses this existential statement is itself a *standard* number. Existential instantiation with a new constant c on the theorem $\exists x T(\ulcorner Y(\text{dot}(x)) \urcorner)$ yields merely $T(\ulcorner Y(\text{dot}(c)) \urcorner)$. There is no guarantee that this constant c denotes a *standard* number. Thus, one *cannot* infer that, for some standard number m , we can prove $T(\ulcorner Y(\underline{m}) \urcorner)$. This is simply not a *deductive* consequence of $\exists x T(\ulcorner Y(\text{dot}(x)) \urcorner)$.⁶

⁵ Briefly, the phenomenon of ω -inconsistency is this. A consistent arithmetical theory θ may have the property that it proves $\sim\varphi(\underline{n})$ for each n , but also proves $\exists x\varphi(x)$. An example is the theory $PA \cup \{\sim G\}$, where G is a Gödel sentence for PA . This phenomenon is closely connected to the fact that no formal first-order theory can 'implicitly define' the natural number structure \mathbb{N} , as there will always be non-standard models available.

⁶ If we want to make *explicit* that our quantifiers range over the natural numbers, we must 'push everything' up a level, to the meta-language, and ask if the Yablo biconditionals are satisfiable on the structure (\mathbb{N}, E) , where \mathbb{N} is the standard model, and E is some choice for the extension of the truth predicate T . Then the *reductio* assumption is not some Yablo biconditional itself, but rather the *meta*-assumption that it is

Here is the mistaken reasoning in question,

This means that there is at least one true sentence in the Yablo list. Let the first such sentence be ' s_i '. (Note that ' i ' is not a variable, but an unknown, particular natural number.) Now consider ' s_i '.

$$Ts_i \Rightarrow \forall k > i, \neg Ts_k. \text{ (2003: 155)}$$

The first statement is mistaken. One cannot *deductively* infer that 'there is at least one true sentence in the Yablo list', for the reason noted above. Furthermore, the parenthetical assertion is incorrect: the ' i ' here is *either a free variable or a new constant*, and *not* (a numeral for) an 'unknown, particular natural number'.

Finally, one *cannot* deductively infer $Ts_i \Rightarrow \forall k > i, \neg Ts_k$. At best, one can infer (in their notation) $s_i \Rightarrow \forall k > i, \neg Ts_k$, unless one explicitly assumes the *uniform* equivalence $Ts_i \Leftrightarrow s_i$, where i is a *free variable*.⁷ For this, one requires the *Uniform* Yablo Disquotation Principle (F), which in our notation is $\forall x(T(\ulcorner Y(\text{dot}(x)) \urcorner) \leftrightarrow Y(x))$.

Priest is quite explicit on the matter:

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 ... and the T-schema applies only to sentences, not things with free variables. (Priest 1997: 237)

In any case, as noted above, the *Local* Yablo Disquotation Principle is in fact consistent. More exactly, PA_Y is a conservative extension of PA . What is needed for any deductive derivation of an inconsistency is a *uniform*

satisfiable on (\mathbf{N}, E) . I.e. $(\mathbf{N}, E) \models Y_n \leftrightarrow \forall x > \underline{n} \sim T(\ulcorner Y(\text{dot}(x)) \urcorner)$. If we further constrain the extension of the truth predicate by demanding that $(\mathbf{N}, E) \models Y_n \leftrightarrow T(\ulcorner Y_n \urcorner)$, for all n , then we do obtain a contradiction. Notice that these two *meta*-assumptions – that all Yablo biconditionals are satisfiable on (\mathbf{N}, E) and that the local Yablo disquotation principle is satisfied on (\mathbf{N}, E) – are equivalent to the (inconsistent) Uniform Homogeneous Yablo Principle above, but now transferred to the metatheory. We then correctly conclude that the set of Yablo biconditionals plus the local Yablo disquotation principle is not satisfiable on (any expansion of) the standard structure \mathbf{N} .

⁷ This is obscured by the notation used by Bueno and Colyvan, a notation which encourages use/mention confusion, as well as confusion about the nature of the index ' i ' on ' s_i '. Note that ' Ts_i ' is syntactically meaningless. T is a monadic predicate, and thus should be followed by a term, either a *variable*, a *function term* or a *name* of a sentence. Usually this name will be a quotation name, with the result $T(\ulcorner s_i \urcorner)$. But there is now a problem with the index ' i '. We want to use this as a variable, since it appears as a variable in the Yablo sentences themselves, and we need to use it as a variable in quantificational reasoning with these sentences. When this is done more precisely, we obtain the formalization given above (where the ' $\text{dot}(x)$ ' construction is necessary).

disquotation principle, the equivalence of the Yablo formula $Y(x)$ and the formula $T(\ulcorner Y(\text{dot}(x)) \urcorner)$, where x is a free variable, not a specific numeral. That is, the uniform disquotation axiom $\forall x(T(\ulcorner Y(\text{dot}(x)) \urcorner) \leftrightarrow Y(x))$ is indispensable in the inconsistency proof. Equivalently, we need the disquotational axiom $\forall x(\text{Sat}(\ulcorner Y(y) \urcorner, x) \leftrightarrow Y(x))$. And this a *uniform* disquotation principle, not a *local* disquotation scheme.

5. Summary

In combination with the *local* disquotation principle, the list of Yablo biconditionals is not formally paradoxical, but rather ω -paradoxical. It is unsatisfiable on (any expansion of) the standard model \mathbb{N} of arithmetic. This is the *most* that any formal proof related to Yablo's paradox can establish using just the *local* disquotation principle for Yablo sentences. Indeed, the list of Yablo biconditionals is satisfiable on *all* non-standard models, and is formally consistent. Bueno and Colyvan are mistaken in insisting that the list of Yablo biconditionals plus the *local* disquotation principle is sufficient for a *formal* derivation of the paradox.

As already noted, and explained in Priest 1997, one *can* reconstruct the paradox (i.e. one can *deduce* an inconsistency from certain assumptions), but this necessarily requires a pair of *uniform* assumptions: the Uniform Fixed-Point Yablo Principle and the Uniform Yablo Disquotation Scheme. And finally, whether one expresses the relevant disquotation principles using a *truth* predicate or a *satisfaction* predicate is immaterial, since these notions are interdefinable.

University of Cambridge
Cambridge CB3 9DA, UK
jjk32@cam.ac.uk

References

- Boolos, G., J. P. Burgess and R. C. Jeffrey. 2002. *Computability and Logic*. Fourth edition. Cambridge: Cambridge University Press.
- Bueno, O. and M. Colyvan. 2003. Paradox without satisfaction. *Analysis* 63: 152–56.
- Kaye, R. 1991. *Models of Peano arithmetic*. Oxford Logic Guides 15. Oxford University Press.
- Ketland, J. 1999. Deflationism and Tarski's paradise. *Mind* 108: 69–94.
- Ketland, J. Forthcoming. Yablo's paradox and ω -inconsistency. *Synthese*.
- Priest, G. 1997. Yablo's paradox. *Analysis* 57: 236–42.
- Yablo, S. 1993. Paradox without self-reference. *Analysis* 53: 251–52.