THE OLD-FASHIONED YABLO PARADOX

Lavinia Picollo
Universidad de Buenos Aires – CONICET
lavipicollo@gmail.com

Abstract

The Yablo Paradox' main interest lies on its prima facie non-circular character, which many have doubted, specially when formulated in an extension of the language of first-order arithmetic. Particularly, Priest (1997) and Cook (2006, forthcoming) provided contentious arguments in favor of circularity. My aims in this note are (i) to show that the notion of circularity involved in the debate so far is defective, (ii) to provide a new sound and useful partial notion of circularity and (iii) to show there is a non-circular formulation of the list in an extension of the language of first-order arithmetic according to the new notion.

KEY WORDS: Yablo Paradox; first-order languages; circularity; Triviality; Fixed point.

Resumen

El interés principal de la Paradoja de Yablo yace en su carácter prima facie no circular, el cual ha sido puesto en duda especialmente con respecto a la formulación de la paradoja en una extensión del lenguaje de la aritmética de primer orden. Particularmente, Priest (1997) y Cook (2006, en prensa) formularon argumentos contenciosos a favor de la circularidad. Los objetivos de esta nota son (i) señalar que la noción de circularidad utilizada hasta el momento en el debate es defectuosa, (ii) ofrecer una nueva noción parcial adecuada y útil de circularidad y (iii) mostrar que existe una formulación no circular de la lista en una extensión del lenguaje de la aritmética de primer orden de acuerdo con la nueva noción.

PALABRAS CLAVE: Paradoja de Yablo; Lenguajes de primer orden; Circularidad; Trivialidad; Punto fijo.

As is well known, the Yablo Paradox' main interest lies on its prima facie non-circular character, which many have doubted, specially when formulating the list in first-order arithmetical systems.[1] Particularly, Priest (1997) and Cook (2006, forthcoming) provided contentious arguments in favor of circularity. My aim in this note is to show there is a non-circular formulation of the list in a first-order arithmetical

[1] By 'arithmetical system' or 'arithmetical theory' I will understand any classical extension of PA or PA itself.

language.[2] In section 1 I show different ways of formalizing the Yablo sequence within such languages. In 2, I first examine Priest's position and make some critical comments, and then analyze Cook's claims and argue that the notion of circularity he embraces is built on the wrong bases and, therefore, fails to capture the most simple intuitions. Section 3 is devoted to provide more accurate bases on which better criteria can rest, develop such a criterion and show its advantages. Finally, in 4 I conclude there is a first-order formulation of the Yablo sequence that is not circular according to the proposed criterion and ponder the value of this result.

1. Formalizing the Yablo list

As Cook (2006, forthcoming) notices, there are at least two traditional ways of allowing first-order formulae to refer to each other: by means of identity statements and by means of biconditionals. In the case of the Liar Sentence, for instance, the first alternative gives us the formula '$\lambda = <\neg T\lambda>$',[3] stating that λ is the name of the sentence that says of itself that it is untrue, which Cook calls 'Old-Fashioned Liar'; while the second leads to '$\lambda \leftrightarrow \neg T<\lambda>$', according to which λ is a formula merely equivalent to the statement that it is itself untrue. This biconditional expression is known as the 'Arithmetic Liar'. The Yablo sentences can be formalized within arithmetical languages by both means as well. To guarantee the existence of the Arithmetic Yablo sequence:

$$(YA_Y) \qquad \{Y(n) \leftrightarrow \forall y(y > n \rightarrow \neg T<Y(dot(y))>) : n \in \omega\}^4$$

we may add a new monadic predicate symbol Y to L_T, the language that obtains by adding to L_{PA} the monadic predicate symbol T for truth, and formulate each biconditional in the sequence; or we could just apply a generalization of the Diagonalization Lemma to the L_T formula '$\forall y(y > x \rightarrow \neg Tz(dot(y)/<x>))$',[5] obtaining the $UFPYP_Y$:

$$(UFPYP_Y) \qquad Y(x) \leftrightarrow \forall y(y > x \rightarrow \neg T<Y(dot(y))>)$$

[2] By 'arithmetical language' I will understand any classical extension of L_{PA}, the language of PA, or L_{PA} itself.

[3] For any formula Φ of a first-order language, <Φ> denotes Φ, possibly via Gödel coding.

[4] I am using the same symbol for numbers and numerals in order to simplify formulae.

[5] The tryadic function symbol x(y/z) represents the primitive recursive function that, applied to the codes x, y, and z of a formula Φ, a term t and a variable v, respectively, gives the code of the formula that obtains by replacing v in Φ with t.

and then derive the set of Yablo biconditionals from it by replacing x for each numeral in the language. Hence, the Arithmetic Yablo can be constructed in L_T. On the contrary, to guarantee the existence of the Old-Fashioned Yablo sequence, the addition of a new monadic predicate name S to the language is needed. By stipulation, one could define S as follows:

(UFPYP_S) $\quad S = <\forall y(y > x \rightarrow \neg TS(dot(y)/<x>))>$

Again, the Yablo sequence is obtained by instantiating the free variable x with each numeral.[6] However, one can directly stipulate the list, instead of this general principle and get:

(YA_S) $\quad \{S(dot(n)/<x>) = <\forall y(y > n \rightarrow \neg TS(dot(y)/<x>))> : n \in \omega\}$

2. Priest's claims and Cook's interpretation

Priest (1997) considers both ways of formalizing the list and focuses on the formal versions of step marked (**) in the introduction to this symposium. As he claims, there are two ways of reading it. First, n could be naturally seen as a variable, to which universal generalization is applied right away. However, in that case, the first part of the reductio must be carried over turning, not just to biconditionals or identity statements in the list, but also to the more general corresponding principles: UFPYP_Y and UFPYP_S, which, of course, are not entailed by the lists themselves. Priest claims these are both circular formulae, for they provide definitions for Y and S in terms of themselves: while the UFPYP_Y states that Y(x) is a weak fixed point of the formula '$\forall y(y > z \rightarrow \neg Tz(dot(y)/<x>))$' —for it just establishes equivalence, but not identity— the UFPYP_S states that S is a strong fixed point of '$<\forall y(y > x \rightarrow \neg Tz(dot(y)/<x>))>(z/<z>)$'. But neither of these principles is merely arithmetical, merely truth-theoretical, or merely arithmetical and truth-theoretical. Thus, they must belong to the core of their corresponding paradoxes. Since its reasonable to call a paradox circular whenever constituted by circular expressions, Priest concludes that the antinomies obtained this way are circular.

[6] Actually, a little move is needed to obtain the Yablo identity statements from the UFPYP_S: since $S = <\forall y(y > x \rightarrow \neg TS(dot(y)))>$, we can get $S(dot(x)/<x>) = <\forall y(y > x \rightarrow \neg TS(dot(y)))>(dot(x)/<x>)$ and then instantiate x with each numeral.

But, secondly, n could also be seen as schematic, standing for any natural number. From this point of view, Yablo's reasoning would not require the UFPYP in any of its versions, but just the numerical instances, though an infinite number of subproofs would be needed, one for each n, and then an application of the ω-rule, which Priest dismisses for non-logical reasons. Notwithstanding, Priest asserts that, even if Yablo actually applied this rule and not the UFPYP, circularity is still present in the original structure of his construction, for it involves a predicate, Y or S, whose instances of application are all defined in terms of (other instances of) itself and, therefore, are ungrounded. According to Priest, Y and S are still fixed points of the corresponding formulae mentioned above.

Even if we concede Priest that both UFPYPs are circular for stating that Y and S are fixed points of other predicates, his reasons do not suffice to do so with the sets of their corresponding numerical instances, YA_Y and YA_S. First, while, as Priest claims, each member of these sets is defined in terms of Y and S, respectively, and this implies that both YA_Y and YA_S are intuitively ungrounded, that should not be enough for circularity, since we could add a new term c to the language, the formulae 'c > n' for each $n \in \omega$ to the two of them and the sentences 'Y(c)' and 'S(dot(c)/<x>) = 0' to each of them, respectively, grounding both and, thus, making the so-called circularity disappear. Since circularity should not vanish by adding new elements, it is not reasonable to claim that it was present before. Second, despite Priest's assertions, if we just consider members of YA_Y or YA_S, it is not clear that Y and S are still fixed points of the same predicates respectively. In order to make that claim the corresponding UFPYP is needed, and its mere numerical instances do not entail it by themselves. As far as it goes, if we reject both UFPYPs and stick to YA_Y or YA_S there is prima facie no reason to claim circularity is involved, though contradiction will not arise in a first-order system.

According to Cook (forthcoming, p. 98), the reasons that allow Priest to claim that the Arithmetic Yablo is circular with or without the $UFPYP_Y$ are that:

> […] the construction involves some sort of expression that turns out to be equivalent to a second expression that 'says' something about the first (or its Gödel code). Whether or not this expression is one of the statements involved in the paradox or a sub-sentential component of such statements would seem to be irrelevant. This, I take it, is the substantial (and correct!) core of Priest's argument in his (1997).

Cook sticks to Priest's alleged criterion whereby an expression is circular if it involves a formula that is a weak fixed point of some predicate. But then notices that in every first-order language containing names for its expressions any formula is a weak fixed point of some other. Thus, he concludes, each sentence of a first-order arithmetical language is circular in this sense, thereby rejecting any possibility of having a non-circular expression, paradoxical or not, within them, including the Old-Fashioned Yablo in any of its versions. This forces him to embrace infinitary languages to get what he believes is a genuine non-circular antinomy. Finally, he claims that is the sort of circularity present both in the Arithmetic Liar and in the Arithmetic Yablo and, hence, neither it has explanatory power, nor is the root cause of paradoxicality, as it is involved in all formulae.

3. A new criterion

The criterion for circularity in first-order languages adopted by Cook (forthcoming) is visibly unsound. In the first place, it trivializes the notion of circularity for, as Cook claims, according to it all sentences become circular. Moreover, if the criterion were conceptually accurate, it could pass on to sentences in the natural language, where any declarative sentence A of English is a weak fixed point of the predicate "'x' is identical to 'x', and A": A is logically equivalent to "'A' is identical to 'A', and A".[7] Hence, every statement of English would be circular as well, which is obviously not the case.

Nonetheless, it is not the criterion itself that should be blamed for trivialization but a more basic notion involved in it, a notion of reference according to which, given two first-order formulae Φ and Ψ, Φ refers to Ψ if Φ is equivalent to another expression containing <Ψ>. Despite the wide acceptance this notion of reference has among specialists (excluding Cook) —particularly when it comes to the Arithmetic Liar and the Gödel sentence— it is not a materially adequate condition for reference, for even interpreting equivalence with the most weak notion, i.e., logical equivalence, it turns out that every expression refers to each other. To see it, the trick from last paragraph could be mimicked in any first-order

[7] It could be argued that classical logic is not English's logic, or even that English has no logic at all. In the first case, a few modifications in A's classically equivalent statement could be introduced to preserve the equivalence even within other logical systems. In the second case, it is hard to see how a notion of circularity could be applied to declarative sentences of English then.

language but, moreover, propositional logic suffices: any first-order formula Φ is logically equivalent to '$\Phi \wedge (<\Psi> = <\Psi> \vee <\Psi> \neq <\Psi>)$'. Hence, the notion of reference on which Cook's criterion lies already entails triviality and must therefore by rejected as base notion for any criterion (and I guess Cook himself would be happy to do so).

As a consequence, I suggest we should look for a new analysis of reference. Old-fashioned formulations come up naturally for this task, since it seems reasonable to assert that, given two first-order formulae Φ and Ψ, Φ refers to Ψ if Φ itself contains $<\Psi>$ and once an identity statement has been established, we are not allowed to add other expressions to the right of the identity symbol, being at risk for triviality as we were when equivalence was involved.

There is at least another intuitive notion of reference between first-order expressions in the literature, according to which formulae refer to each other if they (or equivalent formula) quantify over them. For instance, in L_T '$\forall x(x = x)$' refers to every formula, including itself, but '$\forall x(Tx \rightarrow Px)$' refers only to true sentences. Of course, a notion of circularity could (and should) be built on this relation, as Leitgeb (2002) does. However, it is not the kind of circularity we are analyzing here, for Yablo's sentences seem to avoid it by quantifying just over the ones below them in the sequence, or at least both Priest (1997) and Cook (2006) concede this point.[8] The notion of reference I provide next, and the corresponding criterion for circularity I introduce, only concern mentioning and say nothing about achieving reference or circularity by quantification or other means. Thus, I will use 'm-reference' and 'm-circularity', respectively. Naturally, m-reference and m-circularity entail, correspondingly, reference and circularity simpliciter, but not the other way around.

Our preliminary way of defining reference suits perfectly for sentences such as '$\tau = <T\tau>$', an old-fashioned version of the Truth-Teller, for it entails that τ is self-referential. However, it only accounts for cases of direct reference as this one, leaving aside reference cycles such as the one given by the sentences '$\tau_1 = <T\tau_2>$' and '$\tau_2 = <T\tau_1>$', where τ_1 refers to itself indirectly. Thus, if Φ and Ψ are formulae of a first-order language:[9]

[8] For specific quotes, see Priest (1997, p. 237) and Cook (forthcoming, pp. 72-73). Some objections had been raised by Alberto Moretti (private conversation), since the nth Yablo statement quantifies over sentences whose code is grater than n, and that may include it. Leitgeb (2002) has reformulated Yablo's list to avoid this. Thus, if the reader finds the objection compelling, she could reproduce Leitgeb's trick without changing the main facts of the paper.

[9] Both following definitions belong to Leitgeb (2002, pp. 4-5).

DEFINITION 1: Φ directly m-refers to Ψ if and only if Φ contains a singular term t and $t = <\Psi>$.

DEFINITION 2: Φ m-refers to Ψ if and only if the ordered pair $<\Phi, \Psi>$ belongs to the transitive closure of direct m-reference.

Thus, while τ both directly and indirectly refers to itself, τ_1 and τ_2 refer to themselves only indirectly, and this is intuitively sound. Now we are in a position to introduce a new criterion for circularity founded on our new notion of m-reference. Following the intuition —shared apparently Cook's (forthcoming) own criterion and by many others— that any formula of a first-order theory is circular if it contains an expression that 'says' something about, or refers to, itself, given any two formulae Φ and Ψ of the language of a first-order system Th:

DEFINITION 3: Φ is m-circular if and only if it contains a singular term t such that $t = <\Psi>$ and Ψ m-refers to Ψ according to Φ and Th.[10]

According to this criterion, it is easy to notice that '$\tau = <T\tau>$' comes out circular, but also does, for instance, '$\neg T\tau_1$' if both '$\tau_1 = <T\tau_2>$' and '$\tau_2 = <T\tau_1>$' are entailed by the background theory, since it contains τ_1, the theory implies that '$\tau_1 = <T\tau_2>$' and, as we have already mentioned, '$T\tau_2$' refers to itself.

Definition 3 only applies to single formulae. However, since we are concerned mainly with the circular or non-circular character of the Yablo sequence, we are also interested in evaluating the possibly circular character of sets of expressions. Let Γ be a set of formulae of the language of a first-order system Th:

DEFINITION 4: Γ is m-circular if and only if one of its members contains a singular term t such that $t = <\Psi>$ and Ψ m-refers to Ψ according to Γ and Th.

Thus, if we follow definition 4, $\{\tau_1 = <T\tau_2>, \tau_2 = <T\tau_1>\}$ turns out to be trivially circular but also $\{\tau_1 = <T\tau_2>, \tau_2 = <T\tau_3>\}$ if the background

[10] In a previous version of this note—the one originally sent to Cook—I had defined m-circularity as containing a singular term t such that $t = <\psi>$ and ψ m-refers to itself. As Cook (in this volume) correctly points out, this definition is unsound. Shortly after writing this note but before getting Cook's comments I noticed my mistake and fixed it precisely in the direction Cook suggests in his Definition 3.5.

theory entails that $\tau_3 = \langle T\tau_1 \rangle$. Allowing the background theory (including logic) to play a role in these criteria does not make them trivial. In fact, identities involving names are not logical truths, but stipulations of one system or another. Also, implication within the background theory is restricted to reference patterns, and it must be, contrary to Leitgeb's (2002) claim. If some background theory entails m-circular formulae and we allowed expressions to be m-circular whenever they imply formulae that turn out to be circular according to definition 3, every such expression would be m-circular too. For take a term t denoting an m-circular expression Ψ within Th ($t = \langle\Psi\rangle$ and Ψ refers to itself according to Th); then, any formula entails '$t = t$' by logic alone and, thus, it would be circular. Intuitively, circularity should emerge from the expression or set of expressions called circular, and not just form logic or background theories.

As we have seen, the new criterion provides the right answer for paradigmatic cases as the Old-Fashioned Liar, the Old-Fashioned Truth-Teller and any of their cycles. At the same time, it does not regard every expression as m-circular, since, for instance, '$0 = 0$' does not contain a term that denotes a self-referential formula. It might be the case that m-circularity is neither sufficient nor necessary for paradox, but being not as vacuous as Cook's suggested notion, it could have prima facie an explanatory role in paradoxicality.

4. What about Yablo?

The aim of this paper is to show there is a formulation of the Yablo list that is not circular, at best in the sense of 'circular' we are concerned with. Of course, given that the new criterion seems sound, it will serve as tool for evaluating different formulations. However, let us begin by noticing that the set of formulae that has been called 'Arithmetic Yablo' does not provide a satisfactory formulation of Yablo's original sequence. For while in the last one sentences refer to each other, the Arithmetic Yablo does not entail reference of any kind, since reference cannot be achieved by mere equivalence.

Thus, we are left only with the two mentioned ways of getting the Old-Fashioned Yablo: by the $UFPYP_S$ and just by the set of its instances, YA_S. If, following Priest's (1997) first path, one believes that Yablo's informal reasoning applies the $UFPYP_S$, this principle belongs to the core of the paradox we obtain, as established in section 2. Since the $UFPYP_S$ contains a term, S, such that $UFPYP_S \vdash S = \langle\forall y(y > x \rightarrow \neg TS(dot(y)/\langle x\rangle))\rangle$ and '$\forall y(y > x \rightarrow \neg TS(dot(y)/\langle x\rangle))$' m-refers to '$\forall y(y > x \rightarrow \neg TS(dot(y)/\langle x\rangle))$'

according to the $UFPYP_S$, the paradox is circular: we agree with Priest. However, if we believe that is not Yablo's move, therefore excluding the $UFPYP_S$ and limiting ourselves just to identity statements in the sequence, we do not get a paradox, but the set we are considering, namely, YA_S, is not circular according to our criterion, for non of its members establishes an identity between a term and a formula that mentions it but always expressions below it:[11] the cycle never closes, we disagree with Priest.

Consequently, contrary to Cook's claim, it seems possible to have non-circular expressions in first-order languages, and it also seems possible to have a formulation of the Yablo list among them. This result may appear to be sterile, since that formulation does not entail a contradiction along with reasonable arithmetic and truth-theoretical principles, but it is not entirely such. Both Cook's criterion and ours extend naturally to second-order languages, with similar consequences, respectively. If we add the Yablo identity statements to (full) second-order arithmetic, PA_2, along with the Uniform T-Schema for such sentences we get an unsatisfiable —though consistent— system.[12] Thus, if we consider the impossibility of assigning stable truth values to expressions in the list —witnessed by the lack of (full) models— is enough for paradox, it would not be necessary, as Cook claims, to appeal to infinitary languages to have a non-circular antinomy.

References

Cook, R. (2006), "There Are Non-circular Paradoxes (But Yablos' Isn't One of Them)", *The Monist,* 89, pp. 118-149.

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

Leitgeb, H. (2002), "What is a Self-Referential Sentence? Critical Remarks on the Alleged (Non-)Circularity of Yablo's Paradox", *Logique et Analyse*, 177-178, pp. 3-14.

Picollo, L. (2012), "Yablo's Paradox in Second-Order Languages: Consistency and Unsatisfiabilility", *Studia Logica*. http://dx.doi.org/10.1007/s11225-012-9399-6- Springer Netherlands- 2012-10-P 1-17

Priest, G. (1997), "Yablo's Paradox", *Analysis,* 57, pp. 236-242.

[11] Of course, as long as $n < S(dot(n)/<x>)$ for all $n \in \omega$, which is the case as long as we have a reasonable Gödel coding.

[12] See Picollo (2012) for formal proofs, though they must be adapted to the old-fashioned formulation of the list.