Eliminating Self-Reference from Grelling's and Zwicker's Paradoxes *

José MARTÍNEZ FERNÁNDEZ and Jordi VALOR ABAD

Received: 07.07.2012
Final version: 03.01.2013

BIBLID [0495-4548 (2014) 29: 79; pp. 85-97]
DOI: 10.1387/theoria.6397

ABSTRACT: The goal of this paper is to present Yabloesque versions of Grelling's and Zwicker's paradoxes concerning the notions of "heterological" and "hypergame" respectively. We will offer counterparts of these paradoxes that do not seem to involve self-reference or vicious circularity.

Keywords: paradox; self-reference; circularity; Yablo.

RESUMEN: El objetivo de este artículo es ofrecer versiones de las paradojas de Grelling (sobre el predicado "heterológico") y de Zwicker (sobre el hiperjuego) inspiradas en la paradoja de Yablo. Nuestras versiones de estas paradojas no parecen involucrar ni autorreferencia ni circularidad viciosa.

Palabras clave: paradoja; autorreferencia; circularidad; Yablo.

It is usually assumed that self-reference—be it in the form of strict self-reference (as in 'this sentence is false') or in the form of cross-references (as in 'the following sentence is true', 'the previous sentence is false')—is an essential ingredient in the production of semantic paradoxes. In 1993 Stephen Yablo challenged this widespread view by presenting an example of a set of paradoxical sentences that apparently does not involve any form of self-reference or even of circularity. Yablo's paradox is presented as an infinite list of sentences s_k, with k ranging over natural numbers, with the following definitions:

(s_0) for all $k > 0$, s_k is not true,

(s_1) for all $k > 1$, s_k is not true,

(s_2) for all $k > 2$, s_k is not true, etc.

Suppose that, for some i, s_i is true. Then for all $k > i$, s_k is not true. In particular, for all $k > i+1$, s_k is not true, therefore s_{i+1} is true, a contradiction. So we reject our initial assumption and conclude that, for all i, s_i is not true. In particular, for all $i > 0$, s_i is not true, hence s_0 is true and we have another contradiction. This shows that reasoning about the truth conditions of the sentences on the list generates a paradox even

* We are very grateful to the referees of this paper for comments that led to substantial improvements of the paper. We want also to thank the audience at the VII Conference of the *Sociedad de Lógica, Metodología y Filosofía de la Ciencia de España* in Santiago de Compostela. Work on this paper benefited from support from the Spanish *Ministerio de Economía y Competitividad,* I+D Project FFI2011-25626, Consolider Ingenio Programme (CSD2009-00056) and the European Programme FP7-PEOPLE-ITN-2008 (Project number: 238128). The first author acknowledges also the funding of the *Ikerbasque Fellowship* of the *Ikerbasque Foundation* and the support of the *ILCLI* at the *University of the Basque Country*.

though no circular cross-references are involved, since each sentence talks about the ones that follow it on the list, but not about itself or about the ones that precede it.

Since Yablo published his paradox, many philosophers have found paradoxes concerning notions such as "truth", "set" or "denotation" that do not seem to require any form of self-reference.[1] This led some of them to defend the view that circularity is not essential in formulating semantic or set theoretic paradoxes. Moreover, it has been argued that by substituting infinity for self-reference in the so-called "paradoxes of self-reference" we can always obtain a paradox.[2] This seems to raise two fundamental questions which are still in need of a fully satisfactory answer: (*i*) we want to know whether every paradox of self-reference or circularity has a "Yabloesque" counterpart; and (*ii*) we also want to know whether these paradoxes can really be excluded from the family of paradoxes of self-reference. While it remains unclear what the right answer to the second question is,[3] it seems more and more obvious that the answer to the first question should be affirmative. Any paradox of self-reference appears to have a counterpart involving some infinite series of expressions (denoting phrases, definitions, sentences, etc.) which generate inconsistencies; and most of the notions that enable us to formulate paradoxes of the first kind can be used to devise paradoxes of the second kind. Our purpose here is to back up this thesis by giving two new examples of Yabloesque paradoxes that can be interpreted as counterparts of Grelling's paradox and of Zwicker's Hypergame paradox.[4]

---

[1] See, for instance, non-self-referential versions of set-theoretic paradoxes in Goldstein (1994), of the paradox of the preface and the paradox of the knower in Sorensen (1998), of the Curry paradox in Beall (1999) and of paradoxes of denotation in Uzquiano (2004).

[2] Sorensen (1998) claims that all paradoxes of self-reference have non-self-referential counterparts.

[3] Priest (1997) argued for the thesis that Yablo's paradox involves some kind of circularity. Sorensen 1998 addressed Priest's criticisms and argued for the non circularity of this and similar paradoxes. See also Beall (2001) (a reply to Sorensen in support of Priest's main thesis), Bueno and Colyvan (2003) and Ketland (2004). It is not our intention to engage in this debate here. We will remain neutral as to whether these paradoxes involve some kind of circularity or not.

[4] In talking about 'counterparts', we do not mean to make a substantial claim. We just want to point out that these paradoxes are related to Grelling's and Zwicker's paradoxes in the same way as Yablo's paradox is related to the Liar. Grelling's and Zwicker's paradoxes are, respectively, paradoxes of the concepts of satisfaction and game which clearly involve some form of vicious circularity—and this could be claimed of the Liar with respect to truth—and our paradoxes affect those very concepts in a *prima facie* non circular way—just as it happens with Yablo's paradox and the concept of truth.

We could consider a strong sense of 'counterpart' according to which being a counterpart would involve being "the same" paradox. It is not clear to us that in this sense Yablo's paradox is a counterpart of the Liar paradox. Why should it be more a version of the simple Liar than, say, a version of the cyclical Liar (the following sentence is true/the previous sentence is false)? Are the simple Liar and the cyclical Liar the same paradox? To give criteria for the identity of paradoxes is a substantial problem that would involve an analysis of the structure of paradoxical arguments that, as far as we know, has not been undertaken yet. A first step into what is, we think, a promising direction are recent works by Cook (2004) and Schlenker (2007a, 2007b) that show that self-reference can be eliminated from certain non-quantificational languages in the following sense: we can systematically transform each self-referential sentence into an infinite set of sentences in a quantificational language satisfying these two conditions: (*a*) they are not self-referential; and (*b*) they preserve the truth-value of the

1. Why Grelling's and Zwicker's paradoxes are paradoxes of self-reference

Consider the predicate 'heterological' (Het, for short). According to its definition, we say that a predicate ⌜P⌝ is heterological if and only if it is not true of its own name or, in other words, if and only if ⌜P⌝ is not itself P.[5] We could capture the truth conditions of a sentence in which ⌜Het⌝ is applied to some predicate ⌜P⌝ by means of the biconditional: Het(⌜P⌝) iff ¬P(⌜P⌝). Predicates like 'long' and 'short', for instance, are heterological and not heterological respectively, for 'long' is not long, whereas 'short' is short. Grelling's paradox arises when we consider the truth of Het(⌜Het⌝), that is, when we consider whether 'heterological' is itself heterological, for then we have that Het(⌜Het⌝) iff ¬Het(⌜Het⌝). This paradox can also be viewed as a paradox involving the semantic notion of satisfaction (Sat, for short), a relation which is meant to meet condition (1):

(1) $\forall y$ (Sat(y, ⌜P(x)⌝) iff P(y)).[6]

We can now define ⌜Het(x)⌝ as the predicate satisfying the following condition:

(2) $\forall x$ (Het(x) iff ¬Sat(x, x))

Let h = ⌜Het(x)⌝. From (1) and (2), it follows

(3) $\forall x$ (Sat(x, h) iff ¬Sat(x, x))

Grelling's paradox arises when we consider the following instance of (3): Sat(h, h) iff ¬Sat(h, h).

In 1908 Russell described as paradoxes of 'self-reference' or 'reflexiveness' a collection of well known semantic and set-theoretic antinomies that, with some additions, are still labelled that way nowadays.[7] Grelling's paradox was included among them together with Russell's paradox about the class of all classes not belonging to themselves and the liar paradox among others. The label 'self-reference', however, is rather puzzling in this case. Strictly speaking, there is no self-reference—i.e., no expression referring to itself or sentence talking about itself—involved in Grelling's or Russell's paradoxes, for instance. This label should be understood in a broad or metaphorical way which is far from clear. The term 'reflexiveness'—coined by Russell too—perhaps fares better in describing what is common to all these paradoxes. Many of them seem indeed to arise when we consider whether some object keeps a certain relation to itself: the fact that the liar sentence *says about itself* (and not about any other sentence) that it is not true is crucial in the inference of a contradiction; and considering whether the Russell class *belongs to itself* is what triggers Russell's paradox. Likewise, the possibility of *applying to itself* the predicate 'heterological' is what renders it inconsistent. But

original sentence. These results are important, but they are not directly applicable to paradoxes that involve the definition of predicates or games.

[5] Henceforth, we abbreviate 'if and only if' as 'iff'.

[6] In this paper, we are only concerned with predicates, understood as open formulas with one free variable. Hence we restrict the satisfaction relation to those formulas.

[7] See Russell (1956, 61).

the label 'reflexiveness' (or 'reflexivity') still fails to single out a clear feature bringing all these paradoxes together.[8] Russell made the analysis of reflexivity more precise when he pointed out a trait that all these paradoxes seem to share: they all involve some sort of vicious circularity. He characterizes the notion of vicious circularity by means of a negative principle:

> 'Whatever involves *all* of a collection must not be one of the collection'; or, conversely: 'If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total.' (Russell 1956, 63)

In all these paradoxes, there is always some entity a and some collection Ω (sometimes $a = \Omega$) such that a is a possible member of Ω, but one that can only be defined by appealing to all the members of Ω. This leads to a vicious circle in the definitions of Ω and a since we need *all* the elements of Ω in order to define a, but we need to determine (on the basis of a's definition) whether a is an element of Ω in order to define Ω. In the case of Grelling's paradox, for instance, given a language L that contains a predicate 'Het', we could identify Ω with the extension of 'Het', this is, with the totality of things that do not satisfy themselves; and a would then be the predicate 'Het'. We have here a straightforward case of circularity according to the above description. We need all the elements of Ω (including all the predicates that do not satisfy themselves) in order to ensure that 'Het' is well defined and determines an extension. But 'Het' itself is a possible member of Ω, so we need to determine (on the basis of the definition of 'Het') whether 'Het' belongs to Ω in order to ensure that Ω exists and is well defined.

Building upon previous works by Bertrand Russell,[9] Graham Priest (1994, 2002) has advanced a structural description of the paradoxes of "self-reference". He calls 'Inclosure Schema'[10] the general description of the relevant elements present in all paradoxes of self-reference.[11] The main elements of IS are a collection Ω, which contains all things satisfying some property φ, and a function δ that maps all subsets x of Ω possessing a certain property ψ (also possessed by Ω) into elements of Ω in such a way that $\delta(x)$ always satisfies two conditions: $\delta(x) \in \Omega$ (*closure*) and $\delta(x) \notin x$ (*transcendence*). A contradiction arises when we consider $\delta(\Omega)$.[12] The relationship between Ω and $\delta(\Omega)$—one of its elements according to the "closure" condition—reproduces exactly the kind of circularity banned by Russell, and described above in terms of the relation of mutual dependence between some collection Ω and a, one of its alleged elements

---

[8] Some members of the so called paradoxes of 'self-reference' or 'reflexiveness'—for instance, König's paradox about 'the least non-describable (or non-definable) ordinal number'—do not clearly involve an object that keeps a certain relation to itself.

[9] Specially, Russell (1905).

[10] IS henceforth.

[11] Those, at least, where the concept of negation has some role to play. V.gr., Curry's paradox can only be viewed as an instance of IS if, for any proposition α, we define '$\neg\alpha$' as '$\alpha \rightarrow \bot$', where '$\bot$' stands for a necessarily false proposition.

[12] See Priest (2002, 133–36).

(which becomes δ(Ω) in IS). Hence, it is reasonable to suppose that if a paradox can be regarded as an instance of IS and its generation involves the presence of two crucial structural elements that we can easily identify with Ω and δ(Ω) in IS, then the kind of vicious circularity described by Russell must play some role in the production of the paradox.

Priest shows that the predicate 'heterological' generates an instance of IS (with Ω as described above).[13] Likewise, it can be shown that Zwicker's paradox provides also an instance of IS. This paradox is formulated in terms of games:

Let us consider two-player games, like chess or tic-tac-toe. We will call the players *I* and *II*. A game is well-founded when every play of the game ends after finitely many moves, with no ties. If a game is not well-founded, some plays of the game could go on indefinitely, the result being then undecided. Let us now define a new game called the Supergame.[14] It is a two-player game with the following rules: on the first move player *I* chooses a game, which we will call the subgame. Then player *II* starts playing the subgame and, from that moment on, player *II* of the Supergame becomes player *I* in the subgame and player *I* in the Supergame becomes player *II* in the subgame. The player who wins the subgame will be the winner of the Supergame. The Supergame is obviously not well-founded, since on the first move player *I* can choose a non-well-founded game. The Hypergame is a variation of the Supergame that creates a paradox. The rules of the Hypergame are identical to the rules of the Supergame, except that on the first move player *I* must choose a well-founded game. It is easy to prove that the Hypergame is well-founded: on the first move *I* chooses a well-founded subgame, hence any play of the subgame is bound to end after a finite number of moves. However, since the Hypergame is well-founded, there can be plays in which *I* chooses the Hypergame as a subgame. Then *II* begins to play the Hypergame by choosing a new subgame, and *II* could choose the Hypergame again. They could choose the Hypergame forever, proving that the Hypergame is non-well-founded. Therefore the Hypergame is both well-founded and non-well-founded.

Zwicker's paradox is an instance of IS. It involves a totality Ω, the set of well-founded games, and a particular game δ(Ω), the Hypergame, defined in terms of the elements of Ω.[15] Insofar as the heterological and hypergame paradoxes can be viewed as instances of IS, we can consider them as "paradoxes of self-reference" in this rather broad sense in which 'self-reference' actually alludes to the presence of some sort of vicious circularity not necessarily related to the notion of 'reference'. In the next section, we will offer a version of Grelling's paradox that does not seem to involve circularity of any kind; and in the last section, we will offer a Yabloesque counterpart of Zwicker's paradox.

But before moving on, let us address a possible objection. We have used IS—as Priest and others do—as a criterion for determining when a paradox is a paradox of

[13] See Priest (1994, 31–32; 2002, 145–46).

[14] For the Supergame and the Hypergame paradox we follow the presentation in Barwise and Moss (1996). See also Zwicker (1987).

[15] For details, see Valor Abad (2008, 197–98).

circularity. As a matter of fact, Priest argues that Yablo's paradox is an "inclosure" (an instance of IS). Does that mean, after all, that Yablo's paradox and perhaps other Yabloesque paradoxes involve some form of circularity (or at least that we are claiming so)? Well, if it is true that Yablo's paradox is an inclosure, then we have very good reasons for thinking that it is a paradox of circularity, for **Ω** and **δ**(**Ω**) *clearly violate Russell's principle*. But is Yablo's paradox an inclosure?

When Priest casts this paradox into the mould of IS, he presents it as a paradox concerning the satisfaction relation, not the truth predicate. One of the reasons he gives is that, if we interpret the paradox as one involving the truth predicate, we cannot make sense of the contradiction we reach from the sentences s_i in section 1. He expresses the paradox as follows (where 'T' stands for 'true'):

$Ts_n \quad \Rightarrow \quad \forall k > n, \neg Ts_k \quad (*)$

$\Rightarrow \quad \neg Ts_{n+1}$

But:

$Ts_n \quad \Rightarrow \quad \forall k > n, \neg Ts_k \quad (*)$

$\Rightarrow \quad \forall k > n+1, \neg Ts_k$

$\Rightarrow \quad Ts_{n+1} \quad (**)$

> 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 $\neg Ts_0$). (Priest 1997, 237)

According to Priest, for this proof to be valid, we should justify the lines marked with stars by appealing to the Tarskian Schema (TS). However, this is not possible because all lines where TS should be applied contain a free variable, '*n*', and TS can only be applied to sentences, not to open formulas with free variables. The subscript notation in Yablo's paradox obscures the fact that we need a function '*s*' in order to build, for each number *i*, a name, s_i, of the *i*th Yablo sentence. This function "is defined by specifying each of its values, but each of these is defined with reference to *s*" because *s* appears in the sentences themselves (Priest 1997, 239). The idea is that we cannot understand this paradox without thinking of a function *s* whose values can only be defined or specified by appealing to that very function in a circular (and not just recursive) way. The crucial claim here is that we can only understand Yablo's paradox if it involves a circular predicate *s*. If this claim were wrong, the identification of Yablo's paradox with the inclosure described in Priest 1997 would be ungrounded. But Priest offers a powerful reason in favour of his thesis. He raises a legitimate worry: *how do we know that the Yablo sequence exists*? We can prove by an application of Gödel's diagonal lemma the existence of a formula $s(x)$, satisfying the general principle $\forall x\,(s(x) \leftrightarrow \forall k > x, \neg \mathrm{Sat}(k, \ulcorner s(x) \urcorner))$. The instances $s(n)$ of this formula (for *n* a natural number) can now be interpreted as the Yablo sentences. This principle guarantees the existence of the Yablo sequence, but it forces us to express the sequence by means of a predicate

$s(x)$.[16] The contradiction would now be obtained by appealing to the Satisfaction Schema.[17]

If Priest is right in claiming that this suffices for showing that Yablo's paradox is a paradox of circularity, then virtually *every* Yabloesque paradox is a paradox of circularity, for it seems impossible for us to word those paradoxes or think of them without appealing at some point to a functional expression that is circular in the way just described by Priest. But Priest's claim is controversial. Sorensen (1998, 144-46), for instance, argues that the fact that we cannot find in language or in our thoughts non-circular ways of representing the propositions expressed by the Yablo sentences does not mean that *the propositional content* of those sentences is circular. There could even be, for instance, a sequence of propositions that (*i*) originates a Yabloesque paradox, but (*ii*) cannot be expressed by means of a sequence of recursively specifiable sentences. Moreover, there is no agreement concerning how to understand the terms 'self-reference' or 'circularity'. Leitgeb has argued that in this debate two different notions of self-referentiality and circularity are being used: one of them is based on the intuition that a sentence is self-referential when, roughly speaking, the sentence contains a term that refers to itself; and the second notion is based on the idea that a sentence is self-referential when it is obtained as a fixed point of a syntactic function (just as the Yablo sentences have been obtained). Yablo's sequence is circular only in this second sense, not in the first. If these two notions could be made precise, we would have a way to make compatible both intuitions: the second sense would vindicate the

---

[16] Strictly speaking, the satisfaction predicate is not essential. The predicate $\underline{s}$ required for the formulation of the paradox can also be expressed with the truth predicate, for both predicates are interdefinable. See Ketland (2004, 168; 2005).

[17] One might think that there is no need to resort to a circular predicate and to the satisfaction relation in order to prove that the Yablo sentences are inconsistent. The following reasoning offered by Bueno and Colyvan (2003) may suggest that it is possible to derive a contradiction from these sentences by appealing exclusively to TS:

(1) If we substitute 0 for n in the reasoning advanced by Priest and suppose Ts_0, we infer Ts_1 and $\neg Ts_1$ justifying the starred lines by applications of TS on s_0 and s_1.

(2) We then conclude $\neg Ts_0$ by reductio, which means that $\exists k > 0$, Ts_k.

(3) Now, if we let u be some natural number satisfying the existential claim in (2), then, by substituting u for n in Priest's reasoning, we infer another contradiction: Ts_{u+1} and $\neg Ts_{u+1}$ and, again, we use TS.

Ketland, however, has shown that this reasoning is flawed (we thank an anonymous referee for calling our attention to this fact). If it were correct, it would prove that the Yablo sentences and their corresponding TS biconditionals are jointly inconsistent, but this is not the case. These sentences are indeed ω-inconsistent: if—as it happens in standard models—the only values taken by their variables are natural numbers, then we can draw a contradiction in the way just sketched. However, Jeffrey Ketland has proved that there are non-standard models of arithmetic where the Yablo sentences and their corresponding TS biconditionals are *consistent*. (For details, see Ketland 2005.) Steps (1) and (2) in the above reasoning are correct and show that $\exists k > 0$, Ts_k, but, as Ketland (2004, 170) points out, "one *cannot* infer from this that an object which witnesses this existential statement is itself a *standard* number," as we do in (3). And, if u is *not* a standard number, we can no longer infer Ts_{u+1} and $\neg Ts_{u+1}$. Hence, Bueno and Colyvan's attempt to prove the inconsistency of the Yablo sentences without appealing to the satisfaction relation and to a circular predicate fails.

intuition that we prove the existence of the series of sentences using an intuitively circular device, the first would vindicate the intuition that the Yablo sentences are such that they do not refer to themselves at all.[18] These difficult issues go far beyond the scope of this paper and we will not take a stand on them here.

2. A counterpart of Grelling's paradox

The present paradox arises in an interpreted first-order language L containing an infinite number of names: h_0, h_1, h_2,… and two relational terms: '>' and 'Sat'. The interpretation of the h_i names is given by the following list of predicates, which we will call l:

(h_0) $x > h_0 \rightarrow \forall y\,(y > x \rightarrow \neg \mathrm{Sat}(y, x))$

(h_1) $x > h_1 \rightarrow \forall y\,(y > x \rightarrow \neg \mathrm{Sat}(y, x))$

(h_2) $x > h_2 \rightarrow \forall y\,(y > x \rightarrow \neg \mathrm{Sat}(y, x))$

…

For simplicity, we take the domain of interpretation of L to consist only of the predicates of the list l. The relational term '>' establishes an ordering among those predicates in such a way that, for every two natural numbers n and m, $h_n > h_m$ iff n is greater than m.[19] Finally, 'Sat' is the satisfaction relation: $\ulcorner \mathrm{Sat}(x, y) \urcorner$ is true iff x satisfies y.

As a consequence of the previous semantic stipulations, the satisfaction conditions of the predicates of l can be represented by means of the schema (Sh_i):

(Sh_i) $\forall x\,(\mathrm{Sat}(x, h_i)$ iff $(x > h_i \rightarrow \forall y\,(y > x \rightarrow \neg \mathrm{Sat}(y, x))))$

(Sh_m)—the result of replacing 'i' with the natural number m in the former schema—says that, for every predicate x in the list l, x satisfies the predicate h_m if and only if one of the following three cases obtains:

(*i*) x is identical to h_m;

(*ii*) x precedes h_m in l ($h_m > x$);

(*iii*) x follows h_m in l ($x > h_m$) and no predicate following x in l satisfies x.

We can show that the predicates of l give rise to a paradox. But, in order to do so, we first need to establish a lemma:

(*Lemma*) For every x, y, z, if $x > y$ and $\mathrm{Sat}(x, y)$, then $\mathrm{Sat}(x, z)$.

This lemma says that, if any predicate x of l satisfies any other predicate preceding x in l, then it satisfies *all* predicates of l. It is easy to see that the lemma holds. Take any two predicates of l, h_n and h_m, such that $h_n > h_m$ and assume that $\mathrm{Sat}(h_n, h_m)$. According to (Sh_n), $\mathrm{Sat}(h_n, h_m)$ is equivalent to $h_n > h_m \rightarrow \forall y\,(y > h_n \rightarrow \neg \mathrm{Sat}(y, h_n))$. Given that $h_n > h_m$, we infer (I) by *Modus Ponens*:

[18] Leitgeb argues that the task of giving a precise content to these notions of circularity faces very serious problems. More on this issue can be found in Urbaniak (2009).

[19] We will often use the expressions 'x follows y in l' and 'y precedes x in l' as equivalent to '$x > y$'.

(I) $\forall y\,(y > h_n \rightarrow \neg \mathrm{Sat}(y, h_n))$

Next, we want to show that *all* predicates of l are satisfied by h_n. In order to do so let us divide the predicates of l in two groups: (*a*) predicates preceding h_n in l (at least h_m is in this group); and (*b*) predicates not preceding h_n in l (h_n, h_{n+1}, h_{n+2}, and so on). Now, take an arbitrary predicate h_r of l. Suppose it belongs to group (*b*): $h_r > h_n$ or $h_r = h_n$. We know by (Sh_r) that h_n satisfies h_r iff:

(II) $h_n > h_r \rightarrow \forall y\,(y > h_n \rightarrow \neg \mathrm{Sat}(y, h_n))$

However, h_n trivially satisfies this conditional, for it renders its antecedent false. Hence, h_n satisfies h_r. Suppose now that h_r belongs to group (*a*): $h_n > h_r$. As we saw, h_n satisfies h_r iff (II) is the case. But, clearly, the antecedent of (II) is true and we have already proved the truth of its consequent, namely, (I). Therefore (II) is true and so is $\mathrm{Sat}(h_n, h_r)$. At this point, we have shown that if $h_n > h_m$ and $\mathrm{Sat}(h_n, h_m)$, then $\mathrm{Sat}(h_n, h_r)$. Given that h_m, h_n and h_r were arbitrary predicates of l, we can generalize the result and prove the lemma.

How do we get a paradox now?

Take any two predicates of l: h_n and h_m, such that $h_n > h_m$ and suppose that $\mathrm{Sat}(h_n, h_m)$. By (Sh_m), this is equivalent to a conditional, $h_n > h_m \rightarrow \forall y\,(y > h_n \rightarrow \neg \mathrm{Sat}(y, h_n))$, whose antecedent is true. We infer then (I*):

(I*) $\forall y\,(y > h_n \rightarrow \neg \mathrm{Sat}(y, h_n))$

An instance of (I*) gives us $\neg \mathrm{Sat}(h_{n+1}, h_n)$, which is equivalent in classical logic to (II*) given (Sh_n):

(II*) $h_{n+1} > h_n \wedge \exists y\,(y > h_{n+1} \wedge \mathrm{Sat}(y, h_{n+1}))$

The second conjunct of (II*) says that h_{n+1} is satisfied by *some* predicate following it in l. Let us suppose that h_r meets that condition: $h_r > h_{n+1}$ and $\mathrm{Sat}(h_r, h_{n+1})$. We can prove now a contradiction. By applying our lemma on this assumption, we infer:

$\mathrm{Sat}(h_r, h_n)$

However, from (I*) and the fact that $h_r > h_n$, we can also infer:

$\neg \mathrm{Sat}(h_r, h_n)$

This contradiction offers a *reductio ad absurdum* of our initial assumption: $\mathrm{Sat}(h_n, h_m)$. Given that h_n and h_m were arbitrary predicates such that $h_n > h_m$, we have proved that there cannot be a pair of predicates x, y in l such that $y > x$ and $\mathrm{Sat}(y, x)$ or, equivalently, that for every x, y in l, if $y > x$, then $\neg \mathrm{Sat}(y, x)$. Unfortunately, our present conclusion also leads to a contradiction.

Take two predicates of l, h_n and h_m, such that $h_n > h_m$. Then we have $\neg \mathrm{Sat}(h_n, h_m)$, which, given (S$h_m$), is equivalent in classical logic to:

(II') $h_n > h_m \wedge \exists y\,(y > h_n \wedge \mathrm{Sat}(y, h_n))$

The second conjunct of (II') says that *there is* some predicate y in l such that $y > h_n$ and $\mathrm{Sat}(y, h_n)$. But this contradicts what we just showed, namely, that, *for every* x, y in l, if $y > x$, then $\neg \mathrm{Sat}(y, x)$.

This proves that the satisfaction conditions of the predicates of *l* are paradoxical: whenever we consider, for any two predicates x, y of *l* such that $x > y$, whether x satisfies y, we reach a contradiction both, from $\mathrm{Sat}(x, y)$ and from $\neg\mathrm{Sat}(x, y)$.

Notice that there is an important difference between the predicates in *l* and the predicate *h* (heterological). Grelling's paradox arises when we apply 'heterological' to itself—that is why it was considered a paradox of "self-reference" or "reflexivity". Given the definitions of 'heterological' and 'satisfaction', we have that $\mathrm{Sat}(h, h)$ iff $\mathrm{Het}(h)$ iff $\neg\mathrm{Sat}(h, h)$. Applications of 'heterological' to predicates other than itself are not problematic at all. In the case of the predicates of *l*, on the contrary, all instances of the schema $\mathrm{Sat}(h_i, h_j)$ resulting from replacing 'i' with a natural number are straightforwardly true.[20] The possibility of self-applying one of these predicates is not a source of paradox. Moreover, the problematic cases: those where we consider whether $\mathrm{Sat}(x, y)$ when $x > y$, do not seem to depend on the truth conditions of $\mathrm{Sat}(z, z)$ for any z of *l*.

A second thing to notice is that the paradox we have offered is *different* from (Priest's version of) Yablo's paradox. According to Priest and Ketland, the existence of the Yablo sentences –and their paradoxical status– can only be ensured because it is possible to prove the principle (Y): $\forall x\,(s(x) \leftrightarrow \forall k\,(k > x \rightarrow \neg\mathrm{Sat}(k, \ulcorner s(x) \urcorner))$ by means of Gödel's diagonal lemma. Each instance of (Y) (with 'x' ranging over the natural numbers) can then be identified with one of the Yablo sentences. This offers crucial evidence for the claim that Yablo's paradox involves the satisfaction relation 'Sat' and a circular predicate '$s(x)$'. Here, we have described a Yabloesque paradox involving an infinite sequence of predicates and the notion of satisfaction. One might wonder whether Yablo's paradox (under Priest's description) is really different from ours, since they both seem to involve predicates and the satisfaction relation. That they are different can easily be seen when we proceed to justify the existence of *l*—our sequence of predicates—in the same way as Priest and Ketland proved the existence of the Yablo sentences. The diagonal lemma establishes the following general principle (where the variables range, as before, over natural numbers):

(H) $\forall x, y\,(h(x, y) \leftrightarrow (y > x \rightarrow \forall z\,(z > y \rightarrow \neg\mathrm{Sat}(\langle z, y\rangle, \ulcorner h(x, y) \urcorner))))$.

We can then use (H) in order to define the predicates of *l*, just as we used (Y) in order to define the Yablo sentences. The sentence '$s(n)$' corresponds to the nth Yablo sentence, and the formula '$h(n, y)$' corresponds to the nth predicate of *l*:

(Y*) $s(n) \leftrightarrow \forall k\,(k > n \rightarrow \neg\mathrm{Sat}(k, \ulcorner s(x) \urcorner))$,

(H*) $h(n, y) \leftrightarrow (y > n \rightarrow \forall z\,(z > y \rightarrow \neg\mathrm{Sat}(\langle z, y\rangle, \ulcorner h(x, y) \urcorner)))$.

The key difference between the two paradoxes lies in the fact that Yablo's paradox concerns sentences and the paradox described in this section concerns predicates, which is indeed the most relevant difference we find between the Liar paradox and Grelling's paradox.

---

[20] In general, $\mathrm{Sat}(x, y)$ is always true when $y > x$ or $y = x$. This is due to the fact that, according to our definition of the predicates in *l*, x satisfies y iff a conditional whose antecedent is '$x > y$' is true.

3. A counterpart of Zwicker's paradox

In order to devise a Yabloesque counterpart of the Hypergame paradox, consider an arbitrary non-empty set W of well-founded games. Now we will create a series of new games G_1, G_2 ..., not contained in W, such that a contradiction follows from the specification of their rules. The games G_1, G_2 ... will be called G-games. We want to specify the rules of the G-games in such a way that it is completely clear that there is no vicious circularity involved in a play of any of the G-games.

Game G_i is defined with the following rules. On the first move player *I* chooses either a well-founded game G_j, $j > i$ (if there is any such), or any game in W. The chosen game will be called the subgame. Then they play the subgame, player *II* acting as player *I* in the subgame. The winner of the subgame will also be the winner of the game G_i.

The G-games are not circular in any plausible sense, since the rules of any G_i prevent players from choosing either G_i itself or games which can make G_i itself to be chosen at a later move. Nonetheless it is possible to prove a contradiction. Let us see how.

We will call F the set of G-games which are well-founded. Suppose F is finite. Let i be the highest index of the games in F (when F is empty, take $i = 0$). By the choice of i, all G_j, $j > i$, are not well-founded. Fix any $k > i$ and let us consider how the non-well-founded game G_k could develop. Since for every $j > k$, G_j is not well-founded, on the first move of the game G_k player *I* has to choose a subgame in the set W. Since the subgame of G_k is well-founded, any play of the subgame will end in a finite time, so G_k is well-founded. But G_k was a non-well-founded game. This contradiction shows that F cannot be a finite set. Hence there is a sequence G_{i1}, G_{i2} ... with $i1 < i2 <$ ... and all $G_{ij} \in$ F. When playing game G_{i1}, player *I* could choose G_{i2} on the first move and then player *II* could choose G_{i3} on the second move and so on. This shows that G_{i1} is not well-founded, contradicting that $G_{i1} \in$ F.

We would like to point out that the paradox can be reproduced even if the set W is empty, making a slight modification in the definition of the G-games. In the case where W is not empty, imagine that we well order the G-games and the games in W, putting first the G-games and then the games in W. Then we could describe informally the G-games with the rule: search for well-founded games down the sequence, pick up one (the subgame) and play it, letting the other player begin the subgame. When W is empty, this interpretation would give us the following rules for G_i: *I* searches for a well-founded game G_j, $j > i$. If there is none, *I* loses. Otherwise *I* chooses one of them, say G_k, and *II* begins playing G_k. The winner of G_k will also be the winner of G_i. Considering the definition of all G-games at once, the rules of G_i can be described equivalently as follows: players *I* and *II* must alternatively choose well-founded games G_{i1}, G_{i2} ... with $i < i1 < i2 <$ ... The player that cannot choose any well-founded G-game with a higher index than those previously chosen loses the game G_i.

If the G-games are defined in this way, a contradiction can be found modifying the proof given when W is not empty. Using the same notation introduced there, suppose that F is finite and i is the highest index of the members of F. Then all G_j, $j > i$, are not well-founded. Hence *I* loses on the first move of any play G_j, $j > i$, since it is im-

possible to choose any well-founded game. This proves that games G_j, $j > i$, end on the first move and are well-founded after all. The contradiction implies that F is infinite and a new contradiction follows exactly as in the case of non-empty W. This modified argument shows that only the definition of the G-games is responsible for the paradox, the games in W being irrelevant.

4. Conclusions

In this paper we have introduced two paradoxes concerning the predicates of satisfaction and game which are *prima facie* non-circular and are related to Grelling's and Zwicker's paradoxes in the same way as Yablo's paradox is related to the Liar. We have not taken a stand here as to whether Yablo's paradox is really a paradox of self-reference or not, but any verdict we reach in this respect should also apply to the paradoxes described here. We have suggested that the possibility of casting a paradox into the Inclosure Schema is a relevant criterion for deciding whether it involves some form of circularity or not. However, more work needs to be done in order to characterize the notions of self-reference and circularity if we want to decide whether Yablo's paradox (and related ones) are really paradoxes of self-reference.

REFERENCES

Barwise, J. and Moss, L. 1996. *Vicious Circles. On the Mathematics of Non-Wellfounded Phenomena*. Stanford: CSLI Publications.

Beall, JC. 1999. Completing Sorensen's menu: a non-modal Yabloesque Curry. *Mind* 108: 737-39.

—. 2001. Is Yablo's paradox non-circular? *Analysis* 61: 176-87.

Bueno, O. and M. Colyvan. 2003. Paradox without satisfaction. *Analysis* 63: 152-56.

Cook, R. 2004. Patterns of Paradox. *Journal of Symbolic Logic* 69(3): 767-774.

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

Ketland, J. 2004. Bueno and Colyvan on Yablo's paradox. *Analysis* 64: 165-72.

—. 2005. Yablo's paradox and ω-inconsistency. *Synthese* 145: 295-302.

Leitgeb, H. 2002. What is a self-referential sentence? Critical remarks on the alleged (non-)circularity of Yablo's paradox. *Logique & Analyse* 177–178: 3–14.

Priest, G. 1994. The Structure of the Paradoxes of Self-Reference. *Mind* 103: 25-34.

—. 1997. Yablo's paradox. *Analysis* 57: 236-42.

—. 2002 (2nd edition). *Beyond the Limits of Thought*. Oxford: Oxford University Press.

Russell, B. 1905. On Some Difficulties in the Theory of Transfinite Numbers and Order Types. *Proceedings of the London Mathematical Society,* (series 2) 4: 29-53. Reprinted in D. Lackey, ed. *Essays in Analysis*. London: Allen and Unwin, 1973.

—. 1956. Mathematical Logic as based on the Theory of Types. In *Logic and Knowledge*, edited by R. Ch. Marsh, 59-102. London: Allen & Unwin. Originally published in *American Journal of Mathematics* 30 (1908): 222-62.

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

—. 2007b. How to eliminate self-reference: a précis. *Synthese* 158: 127-38.

Sorensen, R. 1998. Yablo's paradox and kindred infinite liars. *Mind* 107: 137-55.

Urbaniak, R. 2009. Leitgeb, "about," Yablo. *Logique & Analyse* 207: 239–254.

Uzquiano, G. 2004. An infinitary paradox of denotation. *Analysis* 64: 128-31.

Valor Abad, J. 2008. The Inclosure Scheme and the Solution of the Paradoxes of Self-Reference. *Synthese* 160: 183-202.

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

Zwicker, W. 1987. Playing Games with Games: The Hypergame Paradox. *The American Mathematical Monthly* 94: 507-514.

José Martínez Fernández is associate professor (*professor agregat*) at the Department de Lògica, Història i Filosofia de la Ciència at the University of Barcelona and a researcher at the Logos Research Group. His work is focused on philosophical logic, mainly in the use of non-classical logics for the solution of semantic paradoxes.

Address: Departament de Lògica, Història i Filosofia de la Ciència, Facultat de Filosofia, Universitat de Barcelona, Montalegre 4, Barcelona 08001, Spain. E-mail: jose.martinez@ub.edu

Jordi Valor Abad has a PhD in Philosophy from the University of Valencia (Spain), where he currently works as a lecturer at the Department of Logic and Philosophy of Science. His research focuses on different topics in the areas of philosophy of language and philosophy of logic. In this area he is particularly interested in issues concerning the structure of paradoxes of self-reference.

Address: Departament de Lògica i Filosofia de la Ciència, Facultat de Filosofia i Ciències de l'Educació, Universitat de València, Avda. Blasco Ibáñez 30, 46010 València, Spain. E-mail: jordi.valor@uv.es