

# YABLO WITHOUT GÖDEL

Volker Halbach      Shuoying Zhang

5th July 2016

*Abstract.* We prove Yablo's paradox without the diagonal lemma or the recursion theorem. Only a disquotation schema and axioms for a serial and transitive ordering are used in the proof. The consequences for the discussion on whether Yablo's paradox is circular or involves self-reference are evaluated.

## YABLO'S PARADOX IN ARITHMETIC

Yablo (1985, 1993) formulated his paradox as an infinite list of sentences. One way to state the paradox is the following:

- (S<sub>1</sub>)                      For all  $k > 1$ ,  $S_k$  is not true.
- (S<sub>2</sub>)                      For all  $k > 2$ ,  $S_k$  is not true.
- (S<sub>3</sub>)                      For all  $k > 3$ ,  $S_k$  is not true.
- ⋮

To state the paradox in a more rigorous way, we would like to define a function on the natural numbers assigning to each number  $i$  the sentence  $S_i$ . Since the definition of  $S_i$  presupposes the sentences  $S_k$  with  $k > i$  this cannot be done recursively. However, such a function can be defined using the second recursion theorem (see Cantini 2009, p. 987, fn. 242). Alternatively, the Gödel diagonal lemma can be employed and this method has become the standard way of presenting Yablo's paradox in arithmetic since Priest's (1997).

The standard method for presenting the liar paradox in arithmetic relies on the diagonal lemma as well. The whole machinery of arithmetization can be avoided by using a binary satisfaction predicate. The paradox becomes a variant of Russell's paradox. In its form with 'heterological' it's known as Grelling's paradox. Here in this paper we show that also Yablo's paradox can be obtained without any arithmetization and, in particular, without any use of the recursion theorem, the diagonal lemma or related constructions, following the pattern of the 'heterological'-version of the liar paradox.

It has been claimed that Yablo's paradox involves circularity because its proof relies on the Gödel diagonal lemma, the recursion theorem, or the like. One cannot argue for the circularity of our version of the paradox on these grounds as none of these tools are used.

In fact, we obtain the paradox in a very weak theory. It doesn't prove any theorems about expressions or their codes. By proving the paradox in minimal settings we come closer to a better understanding what is required for the paradox. There is not just one single set of minimal assumptions that suffice for the paradox. We make use of a ternary satisfaction predicate, while others rely on a unary truth predicate. In this respect our proof needs a stronger device than the proofs in an arithmetical setting, for instance. By considering various versions of the Yablo paradox with minimal sets of assumptions we hope to further our understanding of what the source of the paradox is.

#### YABLO'S PARADOX WITHOUT ARITHMETIC

We work in a language of predicate logic. Apart from the identity symbol, the language contains a binary predicate symbol  $<$  and a ternary predicate symbol  $\text{Sat}(x, y, z)$ . For each formula  $\phi$  in the language there is a closed term  $\bar{\phi}$  in the language. This can be achieved by adding countably many new constants  $c_1, c_2, \dots$  to the language and then fixing some 1-1-mapping between the set of constants and the set of formulae in the language with all constants. This mapping doesn't play any role in the proof of the paradox; it only helps to motivate our version TS of the T-sentences below. No assumptions about the structure of sentences are used in our proof. We write  $\bar{\phi}$  for the constant for  $\phi$ . The first two variables of the language are  $x$  and  $y$ .

Using an argument very close to the proof of Yablo's paradox, we show that the theory Y given by the following schema and two axioms is inconsistent:

- (TS)  $\forall x \forall y (\text{Sat}(\overline{\phi(x, y)}, x, y) \leftrightarrow \phi(x, y))$   
 (SER)  $\forall x \exists y x < y$   
 (TRANS)  $\forall x \forall y \forall z (x < y \rightarrow (y < z \rightarrow x < z))$

The first axiom schema is a variant of the 'uniform' T-schema.  $\text{Sat}$  stands for *is satisfied* or *is true for*. It is to be read: *For all  $x$  and  $y$ , the formula ' $\phi(x, y)$ ' is satisfied by  $x$  and  $y$  iff  $\phi(x, y)$* . An instance in English would be the following sentence: *'is bigger than' is satisfied by objects  $x$  and  $y$  iff  $x$  is bigger than  $y$* . The variables  $x$  and  $y$  are fixed; the first two variables in alphabetic order could be used. Generally,  $\text{Sat}(z, v, w)$  expresses that  $z$  is true if the free variable  $x$  in  $z$  is assigned  $v$  and  $y$  is assigned  $w$ . Schemata analogous to TS could be added for formulae with more free variables; but they are not needed for the proof. Variable assignments would be more flexible; but here we don't assume any sequence coding and therefore a ternary satisfaction predicate is used instead of a binary predicate that applies to a formula and a variable assignment.

The axioms SER and TRANS state that  $<$  is a serial and transitive relation. Usually the sentences in Yablo's paradox are indexed by numbers. Ketland (2005) showed for an arithmetical setting that only a serial and transitive relation is needed to obtain the paradox.

For the proof we write  $\forall z > y \phi$  for  $\forall z (y < z \rightarrow \phi)$  and  $\bar{\psi}$  for  $\overline{\forall z > y \neg \text{Sat}(x, x, z)}$ . The first line is obtained by instantiating  $\phi(x, y)$  with  $\bar{\psi}$  in TS. The second step is obtained

by universal instantiation.

- (1)  $Y \vdash \forall x \forall y (\text{Sat}(\bar{\psi}, x, y) \leftrightarrow \forall z > y \neg \text{Sat}(x, x, z))$   
(2)  $Y \vdash \forall y (\text{Sat}(\bar{\psi}, \bar{\psi}, y) \leftrightarrow \forall z > y \neg \text{Sat}(\bar{\psi}, \bar{\psi}, z))$

Now we assume  $\text{Sat}(\bar{\psi}, \bar{\psi}, a)$  for arbitrary  $a$  and reason in  $Y$  as follows:

- (3)  $\forall z > a \neg \text{Sat}(\bar{\psi}, \bar{\psi}, z)$  from (2)  
 $\exists v > a \neg \text{Sat}(\bar{\psi}, \bar{\psi}, v)$  by SER  
 $\exists v > a (\forall z > v \neg \text{Sat}(\bar{\psi}, \bar{\psi}, z) \wedge \neg \text{Sat}(\bar{\psi}, \bar{\psi}, v))$  by TRANS and (3)  
 $\exists v > a (\text{Sat}(\bar{\psi}, \bar{\psi}, v) \wedge \neg \text{Sat}(\bar{\psi}, \bar{\psi}, v))$  by (2)

Thus the assumption  $\text{Sat}(\bar{\psi}, \bar{\psi}, a)$  leads to a contradiction in  $Y$ . Therefore we can conclude  $\neg \text{Sat}(\bar{\psi}, \bar{\psi}, a)$  for any  $a$ :

- $Y \vdash \forall z \neg \text{Sat}(\bar{\psi}, \bar{\psi}, z)$   
 $Y \vdash \forall z > a \neg \text{Sat}(\bar{\psi}, \bar{\psi}, z)$  for arbitrary  $a$   
 $Y \vdash \text{Sat}(\bar{\psi}, \bar{\psi}, a)$  by (2)

But  $\text{Sat}(\bar{\psi}, \bar{\psi}, a)$  is refutable in  $Y$ , as shown above. This concludes the proof.

Because the axioms are so weak, the paradox is open to other interpretations. For instance, TS can be read as a comprehension schema for binary relations. In this case it may be preferable to replace the schema TS with a quantified version  $\exists z \forall x \forall y (\text{Sat}(z, x, y) \leftrightarrow \phi(x, y))$ . The paradox follows in the same way.

#### VISSER'S PARADOX

The schema TS by itself is already inconsistent. This can be seen by choosing  $\neg \text{Sat}(x, x, x)$  as  $\phi(x, y)$  in the schema TS and then instantiating both universal quantifiers  $\forall x$  and  $\forall y$  with  $\neg \text{Sat}(x, x, x)$ . The inconsistency follows in the style of the proof of Russell's paradox. The point of our proof in the previous section and proofs of an inconsistency using a Yablo-style argument in an arithmetical setting is that they seem to avoid a kind of circularity that is present in the simpler proofs.

In general, if we claim that a new, more complicated proof avoids an appeal to a certain technique or assumption, we should substantiate that claim by showing that the new proof actually does us to prove the result from weaker assumptions. The proof in the previous section actually requires the additional assumptions SER and TRANS, which are not needed for the simple proof. In this section TS is weakened to a consistent principle that can be shown to be inconsistent with SER and TRANS by an argument similar to the one given above. Obviously, neither a Russell-style argument nor any other proof can be applied to derive an inconsistency of the weakened schema. Visser's (1989) paradox affords exactly that.<sup>1</sup> We go on to show that a proof very close to the one above can be used to prove the inconsistency of the new weaker schema with SER and TRANS.

<sup>1</sup>Drafts of Visser's chapter were circulated much earlier than 1989. By presenting Visser's paradox as a variation of Yablo's I don't mean to imply any claim about priority.

Visser presented the paradox in an arithmetical setting with infinitely many truth predicates indexed by natural numbers. He considered the set of T-sentences  $T_n\phi \leftrightarrow \phi$  where  $\phi$  is a sentence containing only truth predicates with index  $k > n$  and proved that this set is  $\omega$ -inconsistent over arithmetic. Visser's hierarchy is very much like Tarski's hierarchy of truth predicate except that the indexing of the levels of the hierarchy is reversed. Thus  $T_0$  applies to sentences with truth predicates  $T_n$  with  $n > 0$  and so on. Thus the hierarchy is illfounded, that is, it is an infinitely descending hierarchy of languages.

Visser's paradox can be reformulated to fit our framework without arithmetic. In contrast to (Visser 1989) we do not use infinitely many truth or satisfaction predicates. Instead we use a quaternary satisfaction predicate  $Sat_x(y, z, w)$  whose level index  $x$  is a quantifiable variable. The axioms have to be chosen in such a way that  $Sat_x(y, z, w)$  is a satisfaction predicate for formulae where all quantifiers over indices are restricted to objects  $v$  with  $v > y$ . Therefore, if  $v > w$ , the predicate  $Sat_v(y, z, w)$  is 'lower' than  $Sat_w(y, z, w)$  in the hierarchy. The effect of using a quantifiable variable as level index is that we do not only obtain an  $\omega$ -inconsistency as Visser did but rather an outright inconsistency.

An occurrence of a variable is in index position iff it occurs in the 'hierarchy level' position. In  $Sat_x(y, z, w)$ , for instance, exactly  $x$  occurs in index position. The axioms for  $Sat_x(y, z, w)$  are now formulated in such a way that  $Sat_x(y, z, w)$  applies only to formulae  $y$  about level  $v$  of the hierarchy with  $v > x$ . More precisely we use the following axiom instead of TS:

$$(vs) \quad \forall x \forall y (Sat_y(\overline{\phi(x, y)}, x, y) \leftrightarrow \phi(x, y))$$

All occurrences of variables in index position in  $\phi(x, y)$  must be bound. All occurrences of quantifiers  $\forall v$  or  $\exists v$  (for some variable  $v$ ) in  $\phi(x, y)$  that bind some occurrence of the variable  $v$  in index position must be restricted by  $v > y$ .

For instance  $\exists w \exists v > y Sat_y(w, x, y)$  would be an admissible instance of  $\phi(x, y)$ . If we try to run the liar paradox with vs, we would have to instantiate  $\phi(x, y)$  with  $\neg Sat_x(x, x, x)$  or the like; but  $x$  is in index position and thus it needs to be bound with a restricted quantifier. In  $\exists v(v = x \wedge \neg Sat_v(x, x, x))$  (assuming we had also identity in the language) all variables in index position are bound. But only if we restrict it as in  $\exists v > y(v = x \wedge \neg Sat_v(x, x, x))$ , the result is an admissible instance of  $\phi(x, y)$ . The liar paradox, however, is blocked. The reason is that  $\exists v > y(v = x \wedge \neg Sat_v(x, x, x))$  is only about levels  $v$  with  $v > y$  in the hierarchy.

The proof in the previous section still goes through *mutatis mutandis*, if the formula  $\forall z > y \neg Sat_z(x, x, z)$  is chosen as  $\phi(x, y)$ . It shows that SER and TRANS together with vs are inconsistent. In contrast to TS, the schema vs by itself is consistent. As long as  $>$  is well-founded, that is, for any set of objects there is always a  $<$ -maximal element, models of vs can be defined by induction on  $>$ . For the proof of consistency it suffices to assume that  $>$  denotes the empty relation. In this case the restricted quantifiers  $\exists v > y$  and  $\forall v > y$  in the instances  $\phi(x, y)$  of vs are idling.

Visser's result is an extensional version of the formalized Yablo paradox in the following sense. If a paradox is seen as the inconsistency of assumptions that look plausible, then

the proof in the previous section does not give us a new paradox. It's only 'intensionally' different from the simple Russell-liar paradox, because the proof is different. But the Yablo argument in the previous section does not establish a new inconsistency. The inconsistency of *vs* with *SER* and *TRANS*, however, cannot be obtained by the simple argument, because *vs* by itself is consistent. Therefore we have a new inconsistency and a new paradox.

#### CIRCULARITY

By proving paradoxes from very weak assumption one can hope to reveal what is really needed to arrive at a contradiction and what the source of paradox is. Arithmetic provides us with very strong tools that are not required to arrive at the paradox. In particular, it provides us with the diagonal lemma and related results that are not needed for our version of the paradox, as the proof above shows.

Most formal presentations of Yablo's paradox rely on an arithmetical framework with an additional *unary* truth predicate. Other presentations of Yablo's paradox with a satisfaction predicate such as (Cook 2014) or (Eldridge-Smith 2015) still rely on some theory of syntax or Gödel codes; our theory *Y* doesn't. In contrast to those theories, the axioms used in the present paper do not tell us much about the objects named by the constants  $\bar{\phi}$ . In an arithmetical setting much stronger assumptions are made that may be suspected of contributing to the paradox.

However, there is a price to pay for the elimination of arithmetical or syntactic assumptions: The proof of Yablo's paradox requires a ternary satisfaction predicate, and the proof of Visser's paradox requires even a quaternary predicate. In the arithmetical case, as has been mentioned, a unary truth predicates suffices.

In the present paper we do not advocate a particular analysis of circularity or self-reference. We only would like to explain in which sense Yablo's and Visser's paradox in our setting are not circular or self-referential.

First, the paradox doesn't involve any term that denotes a formula in which the same formula occurs. Thus it isn't self-referential in the sense of what Halbach and Visser (2014) call the Kreisel–Henkin criterion. Secondly, our proofs do not rely on any arithmetical or syntactic techniques – such as recursion theorem or Gödel's diagonal lemma – that are often seen as involving self-reference. The axioms used here are too weak to prove such sophisticated lemmata. Thirdly, in Visser's paradox the satisfaction predicate of level *y* is applied only to lower levels, that is, to formulae in which the level index of the satisfaction predicate is restricted to lower levels in the hierarchy. We could even add the additional axioms expressing that *<* is a linear order to rule out loops in *<*. We exploit Ketland's (2005) observation that only transitivity and seriality are required. If we assume that there are loops, that is, if we assume in addition also  $\exists x \exists y (x < y \wedge y < x)$ , then  $\exists x x < x$  follows by transitivity. In this case the contradiction follows directly from (2) and Yablo's and also Visser's paradox morph into the simple liar–Russell-paradox. Thus the assumption that *<* is circular considerably shortens the proof; but our proof does not need the circularity of *<*. However, we cannot dispense with seriality in the proof.

The proof needs the assumption that  $>$  is illfounded, that is, we need a nonempty set of objects that does not have a  $<$ -maximal object. Whether  $>$  is illfounded because it is circular or because it is an infinite linear ordering without end point doesn't matter. This is in line with other attempts – such as (Halbach et al. 2003) – to argue that it's not just circularity that is a necessary ingredient for the semantic paradoxes, but more generally illfoundedness.

#### BIBLIOGRAPHY

- Cantini, Andrea (2009), Paradoxes, self-reference and truth in the 20th century, in D. M. Gabbay and J. Woods, eds, 'Logic from Russell to Church', Vol. 5 of *The Handbook of the History of Logic*, Elsevier, Amsterdam, pp. 875–1013.
- Cook, Roy (2014), *The Yablo Paradox: An Essay on Circularity*, Oxford University Press, Oxford.
- Eldridge-Smith, Peter (2015), 'Two paradoxes of satisfaction', *Mind* 124, 85–119.
- Halbach, Volker and Albert Visser (2014), 'Self-Reference in Arithmetic I', *Review of Symbolic Logic* 7, 671–691.
- Halbach, Volker, Hannes Leitgeb and Philip Welch (2003), 'Possible Worlds Semantics For Modal Notions Conceived As Predicates', *Journal of Philosophical Logic* 32, 179–223.
- Ketland, Jeffrey (2005), 'Yablo's paradox and  $\omega$ -inconsistency', *Synthese* 145, 295–302.
- Priest, Graham (1997), 'Yablo's paradox', *Analysis* 57, 236–242.
- Visser, Albert (1989), Semantics and the liar paradox, in D. Gabbay and F. Günthner, eds, 'Handbook of Philosophical Logic', Vol. 4, Reidel, Dordrecht, pp. 617–706.
- Yablo, Stephen (1985), 'Truth and Reflection', *Journal of Philosophical Logic* 14, 297–349.
- Yablo, Stephen (1993), 'Paradox Without Self-Reference', *Analysis* 53, 251–252.

VOLKER.HALBACH@NEW.OX.AC.UK  
SHUOYINGSYZHANG@GMAIL.COM