Yablo sequences in truth theories

Cezary Cieśliński

The University of Warsaw, Institute of Philosophy, Warsaw, Poland

Abstract. We investigate the properties of Yablo sentences and formulas in theories of truth. Questions concerning provability of Yablo sentences in various truth systems, their provable equivalence, and their equivalence to the statements of their own untruth are discussed and answered.

Key words: truth, Yablo's paradox, omega-liar

1 Introduction

In 1993 Stephen Yablo presented a new paradox, belonging to the liar-type group, but, as it was claimed, importantly different from the liar (see [12], but cf. also [11] for a similar reasoning). Let $T(x)$ abbreviate "x is true". Consider an infinite sequence of sentences $Y_0, Y_1, Y_2, \ldots$ such that

$$\begin{aligned} Y_0 \text{ states}: &\quad \forall z > 0\ \neg T(Y_z), \\ Y_1 \text{ states}: &\quad \forall z > 1\ \neg T(Y_z), \\ Y_2 \text{ states}: &\quad \forall z > 2\ \neg T(Y_z), \\ &\vdots \end{aligned}$$

In effect Y_n states that all sentences which appear after stage n in the series are not true. Assume now that Y_k is true. Then for any $i > k$, Y_i is not true, and in particular Y_{k+1} is not true. But also for any $i > k + 1$, Y_i is not true, so Y_{k+1}, therefore Y_{k+1} is true after all, which is impossible. Since the reasoning goes for an arbitrary k, we know at this point that all Y_k-s are not true. Therefore Y_0 must be true and in this way we obtain a contradiction.

The above reasoning has been presented by Stephen Yablo as a "paradox without self-reference": apparently it involves no direct or indirect self-referential loops, since on the face of it all the Y_n-s say something only about sentences which appear *later* in the sequence (i.e. after stage n). The question whether the paradox is really self-reference free has been much debated in recent literature.[1] However, the possibility of giving here a general, unanimous verdict seems very problematic: the main obstacle is that at present we do not seem to have a clear

[1] In particular, Sorensen in [9] defends the non-circularity of Yablo's paradox; but see also [8] and [1] for the criticism of the non-circularity claim.

concept of self-reference or (more generally) of 'aboutness', against which the issue of self-referentiality of Yablo sentences could be decided.[2]

In view of that, I will avoid here the notion of self-referentiality, adopting instead a different approach. The central notion will be that of *provability* in various systems of truth; the plan is to consider the following main questions: (1) which Yablo sentences are provable/disprovable in a given truth theory? (2) are all Yablo sentences provably equivalent in a given theory? (3) are Yablo sentences like the liars - are they (provably in a given theory) equivalent to the statements of their own untruth? (4) to what extent does the answer to (1)–(3) depend on our choice of a Yablo formula $Y(x)$?

Question (1) is perhaps the most basic one – we want to know after all whether a given theory settles in any way the issue of Yablo sentences. Questions (2) and (3) are motivated by discussions of (non)self-referentiality of Yablo sentences. While the notion of a self-referential sentence remains vague, one can use precise, formal tools in order to investigate whether all Yablo sentences are one and the same *up to provability* in a given theory,[3] and also whether they all fit into (again, up to provability) the familiar Liar-type pattern. Question (4) relates to the fact that in general it is possible to obtain sentences with quite different properties satisfying one and the same formal constraint. We ask in effect if the Yablo condition by itself (corresponding roughly to a general specification of what is stated by each Yablo sentence) is enough to determine the answers to (1) – (3).[4]

2 Preliminaries

We start with introducing the notions of a Yablo formula and a Yablo sentence with respect to a theory S. In what follows we will always assume that S is a theory formulated in the language L_T, specified as the language of first order arithmetic extended with a one place predicate $T(x)$. We will use Feferman's dot notation, so e.g. if φ is a formula with one free variable, the expression "$\forall x T(\ulcorner \varphi(\dot{x}) \urcorner)$" gets a reading: "for all natural numbers x, the result of substituting a numeral denoting x for a variable free in φ is true". In practice if there is no danger of ambiguity, we will often suppress both the dots and the square corners.

---

[2] This observation was made by Leitgeb. After analysing various unsuccessful attempts to express the notion of self-referentiality, he even voices the suspicion that "the talk of self-referentiality is to be banished from scientific contexts" (see [7], p. 13; but see also [10] for a defence of this notion).

[3] The real issue concerns implications "$Y(n) \to Y(k)$" for $n > k$, since for $n < k$ the implication is trivial given sufficiently strong background theory.

[4] A similar approach was adopted in [3] – a paper devoted to the analysis of Yablo's reasoning with various provability predicates substituted for truth. One of the main issues is then which variants of Yablo sentences are provably equivalent over a background arithmetical theory.

Definition 1 *Let S be a theory in the language L_T. We say that $Y(x)$ is a Yablo formula in S iff it satisfies (provably in S) the Yablo condition, i.e. iff $S \vdash \forall x[Y(x) \equiv \forall z > x \neg T(\ulcorner Y(\dot{z}) \urcorner)]$. Yablo sentences are obtained by substituting numerals for x in $Y(x)$.*

Using familiar diagonal techniques, it is easy to prove the existence of Yablo formulas for all theories extending Robinson's arithmetic.[5]

Theorem 2. *For every theory S in L_T extending Robinson's arithmetic, there is a Yablo formula in S.*

In the proof we employ the diagonal lemma in the following form:

Lemma 3 *Let S be a theory in L_T extending Robinson's arithmetic. Then for every $\varphi(x, y) \in L_T$ there is a formula $\psi(x)$ such that:*

$$S \vdash \psi(x) \equiv \varphi(x, \ulcorner \psi(x) \urcorner)$$

The proof mimics the usual proof of the diagonal lemma for formulas with one free variable.

With the lemma at hand, the argument for Theorem 2 proceeds as follows.

Proof. Fix:

$$\varphi(x, y) := \forall z > x \neg T(sub(y, name(z))).$$

with "$sub(y, s)$" representing the substitution function (which produces the result of substituting s for a free variable in y) and "$name(x)$" representing a function which for an argument x produces as a value a numeral denoting x.[6] By the diagonal lemma, take $Y(x)$ such that:

$$S \vdash Y(x) \equiv \forall z > x \neg T(sub(\ulcorner Y(x) \urcorner, name(z))).$$

Then the formula $Y(x)$ as constructed above is a Yablo formula in S.

Further properties of $Y(x)$ will depend on the choice of S – in particular, on the axioms governing the use of the predicate T. To clear the ground, consider for a start the theory **PAT**, which is obtained from **PA** by extending the language with a new predicate "T". This predicate will be permitted to appear in formulas substituted for schematic axioms of **PA**.[7] Since by Theorem 2 Yablo formulas in **PAT** do exist, one can consider their properties. For **PAT** the following result holds.

[5] This method of constructing Yablo sequences was employed by Priest, see [8]; cf. also Ketland's paper [5].

[6] Strictly speaking, in the context of arithmetic with addition and multiplication, both expressions (i.e. *sub* and *name*) should be treated not as function symbols, but as arithmetical formulas representing appropriate functions on natural numbers.

[7] As defined, **PAT** is not really a theory of truth, with "T" being just a new predicate, without any substance to it, but we find it useful to consider it as a borderline case.

Fact 4 *Let $Y(x)$ be a Yablo formula in* **PAT**. *Then:*

(a) $\mathbf{PAT} \nvdash \exists x Y(x)$
(b) $\mathbf{PAT} \nvdash \exists x \neg Y(x)$
(c) If $Y(x)$ contains a free variable x, then for all natural numbers n and k, if $n > k$, then $\mathbf{PAT} \nvdash Y(n) \rightarrow Y(k)$

Proof. Since T in **PAT** functions just as a new predicate, $\mathbf{PAT} \nvdash \exists x T(x)$ and also $\mathbf{PAT} \nvdash \exists x \neg T(x)$, therefore both (a) and (b) follow trivially. For (c), assume that $Y(x)$ contains a free variable x. Then for every n and k, if $n \neq k$, then $\ulcorner Y(k) \urcorner \neq \ulcorner Y(n) \urcorner$. Consider a model (N, T) obtained by expanding the standard model N with the set $T = \{Y(n)\}$. Obviously $(N, T) \models \mathbf{PAT}$ and since $n > k$, we have: $(N, T) \models Y(n)$; $(N, T) \not\models Y(k)$.

To sum it up: Yablo sentences are neither provable, nor disprovable in **PAT**; they are also not provably equivalent in this theory (assuming that they are different). In fact $Y(0), Y(1)...$ is a sequence of weaker and weaker sentences independent from **PAT**. It's also worth noting that Fact 4 is obtained independently of our choice of the formula $Y(x)$, as long as (for condition (c)) it contains a free variable x.

The next sections contain a discussion of the status of Yablo sentences in two truth theories: Friedman-Sheard system **FS** and Kripke-Feferman theory **KF**.

3 The theory FS

We proceed now to the discussion of the Friedman-Sheard system **FS**, which is obtained by adding to **PAT** compositional truth axioms for negation, binary connectives and quantifiers, together with the rules of necessitation (NEC) and conecessitation (CONEC). We will denote by $\mathbf{FS}^-$ a theory just like **FS**, but with induction restricted to arithmetical formulas only. In effect **FS** is defined as the system extending **PAT** with the following truth-theoretic axioms and rules (Tm^c is the set of constant terms and $Sent_T$ denotes the set of sentences of L_T):

- $\forall s, t \in Tm^c \big(T(s{=}t) \equiv val(s){=}val(t) \big)$
- $\forall x \, \big(\mathrm{Sent}_T(x) \rightarrow (T\neg x \equiv \neg Tx) \big)$
- $\forall x \, \forall y \, \big(\mathrm{Sent}_T(x \wedge y) \rightarrow (T(x \wedge y) \equiv (Tx \wedge Ty)) \big)$
- $\forall x \, \forall y \, \big(\mathrm{Sent}_T(x \vee y) \rightarrow (T(x \vee y) \equiv (Tx \vee Ty)) \big)$
- $\forall v \, \forall x \, \big(\mathrm{Sent}_T(\forall v x) \rightarrow (T(\forall v x) \equiv \forall t \; T(x(t/v))) \big)$
- $\forall v \, \forall x \, \big(\mathrm{Sent}_T(\exists v x) \rightarrow (T(\exists v x) \equiv \exists t \; T(x(t/v))) \big)$

Additional rules of inference are:

$$NEC \quad \frac{\phi}{T\phi} \qquad\qquad \frac{T\phi}{\phi} \quad CONEC$$

For more information about **FS** we refer the reader to [4], where both semantics and proof theory of this system is discussed.

As it turns out, results concerning Yablo sentences in **FS** do not depend on the choice of a Yablo formula $Y(x)$. Let $Y(x)$ be an arbitrary Yablo formula in $\mathbf{FS}^-$ (analogously for full **FS**). We start with the following observation.

Fact 5 $\mathbf{FS}^- \vdash \forall xz[x < z \to (Y(x) \to Y(z))]$

The proof is immediate, from the assumption that $Y(x)$ is a Yablo formula in $\mathbf{FS}^-$.

Now we will show, that all Yablo sentences are provably equivalent in $\mathbf{FS}^-$. In fact a *uniform* equivalence of Yablo sentences is a theorem of $\mathbf{FS}^-$:

Theorem 6. $\mathbf{FS}^- \vdash \forall xz[Y(x) \equiv Y(z)]$.

Proof. Working in $\mathbf{FS}^-$, fix x and z. Assume (wlog) that $x < z$. Then we know (Fact 5) that $Y(x) \to Y(z)$. For the opposite implication, assume $Y(z)$, i.e. $\forall s > z \neg T(Y(s))$. For an indirect proof, assume also $\neg Y(x)$, i.e. $\exists s > x T(Y(s))$. Therefore $\exists s \leq z T(Y(s))$; fix such an s. Since $Y(z)$, we have also: $\neg T(Y(z+1))$. By applying NEC and compositional axioms to Fact 5, we obtain (as a theorem of $\mathbf{FS}^-$): $\forall xz[x < z \to (T(Y(x)) \to T(Y(z)))]$. Since $s < z + 1$ and $T(Y(s))$, we get: $T(Y(z + 1))$, which is a contradiction ending the proof.

Now we present the following two corollaries.

Corollary 7 $\mathbf{FS}^- \vdash \forall xz[T(Y(x)) \equiv T(Y(z))]$.

The proof is immediate, by applying NEC and compositional axioms to Theorem 6. We have also:

Corollary 8 $\mathbf{FS}^- \vdash \forall x[Y(x) \equiv \neg T(Y(x))]$.

Proof. From left to right, the assumption $Y(x)$ gives us $\neg T(Y(x+1))$, so $\neg T(Y(x))$ by Corollary 7. For the opposite implication, assuming $\neg T(Y(x))$ we obtain $\forall z \neg T(Y(z))$ by Corollary 7; therefore $\forall z > x \neg T(Y(z))$, which gives us $Y(x)$.

Corollary 8 shows that in **FS** each Yablo sentence is a liar - it expresses (up to a provable equivalence) its own untruth. The corollary states that this insight can be proved in $\mathbf{FS}^-$ in a uniform manner. Finally we obtain:

Fact 9 *If* **FS** *is consistent, then:*

(a) $\mathbf{FS} \nvdash \exists x Y(x)$ *(b)* $\mathbf{FS} \nvdash \exists x \neg Y(x)$

Proof. For (a), assume that $\mathbf{FS} \vdash \exists x Y(x)$, therefore by Theorem 6 $\mathbf{FS} \vdash \forall x Y(x)$, so in particular $\mathbf{FS} \vdash Y(0)$. An application of NEC and the compositional axiom for general quantifier gives us: $\mathbf{FS} \vdash \forall x T(Y(x))$, so $\mathbf{FS} \vdash T(Y(1))$, but also $\mathbf{FS} \vdash \neg T(Y(1))$ (because $Y(0)$ is provable in **FS**), contradicting the consistency of **FS**.
For (b), assume that $\mathbf{FS} \vdash \exists x \neg Y(x)$, therefore by Theorem 6 $\mathbf{FS} \vdash \forall x \neg Y(x)$. Applying NEC and compositional axioms, we obtain: $\mathbf{FS} \vdash \forall x \neg T(Y(x))$. But then $\mathbf{FS} \vdash \forall x Y(x)$, which together with the first assumption leads to the conclusion that **FS** is inconsistent.

4 The theory KF

We proceed now to the Kripke-Feferman theory, denoted as **KF**. The truth theoretic axioms are listed below. In what follows they will be denoted as KF1-KF13.

(1) $\forall s \forall t \left(T(s = t) \equiv val(s) = val(t) \right)$
(2) $\forall s \forall t \left(T(\neg s = t) \equiv val(s) \neq val(t) \right)$
(3) $\forall x \left(\mathrm{Sent}_T(x) \rightarrow (T(\neg\neg x) \equiv Tx) \right)$
(4) $\forall x \forall y \left(\mathrm{Sent}_T(x \wedge y) \rightarrow (T(x \wedge y) \equiv Tx \wedge Ty) \right)$
(5) $\forall x \forall y \left(\mathrm{Sent}_T(x \wedge y) \rightarrow (T\neg(x \wedge y) \equiv T\neg x \vee T\neg y) \right)$
(6)-(7) Similarly for disjunction
(8) $\forall v \forall x \left(\mathrm{Sent}_T(\forall v x) \rightarrow (T(\forall v x) \equiv \forall t\, T(x(t/v))) \right)$
(9) $\forall v \forall x \left(\mathrm{Sent}_T(\forall v x) \rightarrow (T(\neg\forall v x) \equiv \exists t\, T(\neg x(t/v))) \right)$
(10)-(11) Similarly for the existential quantifier
(12) $\forall t \left(T(Tt) \equiv T(val(t)) \right)$
(13) $\forall t \left(T\neg Tt \equiv (T(\neg val(t)) \vee \neg\, \mathrm{Sent}_T(val(t))) \right)$

When discussing **KF**, two additional axioms are often introduced:

CONS $\quad \forall x \left(\mathrm{Sent}_T(x) \rightarrow \neg(Tx \wedge T\neg x) \right)$

COMPL $\quad \forall x \left(\mathrm{Sent}_T(x) \rightarrow (Tx \vee T\neg x) \right)$

However, we will denote as **KF** the theory with just the axioms KF1-KF13 added to **PAT**. Whenever we discuss a theory with Cons or Compl, we are going to stipulate it explicitly.

In order to characterize the behaviour of Yablo sentences in **KF**, we will need some basic facts about this theory.

4.1 Basic properties of KF

The presentation in this section relies heavily on Cantini's paper [2]; the modifications are introduced in order to handle our specific choice of axiomatization for **KF**.

KF has been proposed as a formalization of Kripkean notion of truth, based on strong Kleene evaluation scheme (see [6]). In Kripke's fixed point construction, truth is interpreted as a partial predicate – its interpretation is given by a pair of sets T^+, T^- called the extension and the antiextension. Given a classical model M of Peano arithmetic, we will consider structures $\mathcal{M} = (M, T^+, T^-)$, with T^+ and T^- being the subsets of the domain of M – such structures are called *partial models* for the language L_T (we assume that only the predicate $T(x)$ is partially interpreted; arithmetical expressions are interpreted classically.) For partial models a satisfaction relation can be defined in the following way (the subscript in "$\models_{sk}$" is for "strong Kleene"):[8]

[8] For the purposes of Definition 10, it is convenient to extend L_T to the language of the model M, i.e. we add constants for all elements of M. In effect for every $a \in M$, $\varphi(a)$ is a formula (or a sentence) of the extended language.

Definition 10

- $\mathcal{M} \models_{sk} s = t$ *iff* $val(s) = val(t)$*; similarly for negated identities.*
- $\mathcal{M} \models_{sk} Tt$ *iff* $val(t) \in T^+$.
- $\mathcal{M} \models_{sk} \neg Tt$ *iff* $val(t) \in T^-$ *or* $\neg Sent(val(t))$.
- $\mathcal{M} \models_{sk} \neg\neg\varphi$ *iff* $\mathcal{M} \models_{sk} \varphi$.
- $\mathcal{M} \models_{sk} \varphi \wedge \psi$ *iff* $\mathcal{M} \models_{sk} \varphi$ *and* $\mathcal{M} \models_{sk} \psi$.
- $\mathcal{M} \models_{sk} \neg(\varphi \wedge \psi)$ *iff* $\mathcal{M} \models_{sk} \neg\varphi$ *or* $\mathcal{M} \models_{sk} \neg\psi$.
- *Similarly for disjunction and its negation.*
- $\mathcal{M} \models_{sk} \forall x\varphi(x)$ *iff for all* $a \in M$ $\mathcal{M} \models_{sk} \varphi(a)$.
- $\mathcal{M} \models_{sk} \neg\forall x\varphi(x)$ *iff for some* $a \in M$ $\mathcal{M} \models_{sk} \neg\varphi(a)$.
- *Similarly for the existential quantifier.*

Since **KF** is a classical theory, its models will be two valued, not partial. However, each model of **KF** can be turned into a partial model with some nice properties.

Definition 11 *For* $(M, T) \models KF$*, we denote:*

- $T^+ = T$
- $T^- = \{z : \neg z \in T^+\}$
- $M^* = (M, T^+, T^-)$

It turns out that M^*, as characterized by Definition 11, satisfies the following:

Theorem 12. *If* $(M, T) \models KF$*, then* $\forall\varphi \in L_T$ $[M^* \models_{sk} \varphi$ *iff* $M^* \models_{sk} T(\varphi)]$.

Idea of the proof. The proof is by induction on positive complexity of φ (see [4], p. 205).[9] For sentence of positive complexity 0 e.g. of the form $\neg T(t)$ we have: $M^* \models_{sk} \neg T(t)$ iff $val(t) \in T^- \vee \neg Sent(val(t))$ iff $\neg val(t) \in T^+ \vee \neg Sent(val(t))$ iff $(M, T) \models T(\neg t) \vee \neg Sent(t)$ iff $(M, T) \models T(\neg T(t))$ iff $\neg T(t) \in T^+$ iff $M^* \models_{sk} T(\neg T(t))$. The rest follows by induction.

Adding Compl or Cons to **KF** produces a theory which is truth-theoretically (although not arithmetically) stronger than **KF**. It can be shown that both directions of the uniform T-schema (i.e. "$\forall x_1...x_n[T(\varphi(x_1...x_n)) \equiv \varphi(x_1...x_n)]$") are provable in theories with Cons and Compl respectively.

Fact 13 *For every* $\varphi(x_1...x_n)$*:*

(a) $KF + \text{Cons} \vdash \forall x_1...x_n[T(\varphi(x_1...x_n)) \rightarrow \varphi(x_1...x_n)]$
(b) $KF + \text{Compl} \vdash \forall x_1...x_n[\varphi(x_1...x_n)) \rightarrow T(\varphi(x_1...x_n))]$

[9] Roughly, the idea is to define the notion of a complexity of a formula in such a way as to guarantee that: atomic and negated atomic formulas have the complexity 0; conjunctions, disjunctions and quantified sentences have the level of complexity greater by one than their disjuncts/conjuncts/formulas after the quantifier; the same for negated conjunctions/disjunctions/quantified sentences; double negation increases the level of complexity by one.

Accordingly, we can't extend **KF** consistently with both Cons and Compl (the full T-schema is known to be inconsistent); it is possible however to add consistently each of this axioms separately.

Idea of the proof. The fact is proved by induction on positive complexity of L_T-formulas. We show only the parts where Cons and Compl are used. This happens in the case when $\varphi := \neg T(x)$. Then we argue as follows.

(a) Working in $KF+$Cons, assume $T(\neg T(a))$, assume also $T(a)$. Then by KF12, $T(T(a))$, which contradicts Cons.
(b) Working in $KF +$ Compl, assume $\neg T(a)$, assume also $\neg T(\neg T(a))$. Then by Compl, $T(T(a))$, so by $KF12$, $T(a)$ - a contradiction.

From Fact 13 the following conclusion about the liar sentence L can be very easily obtained:[10]

Corollary 14 *$KF +$ Cons $\vdash L$; $KF +$ Compl $\vdash \neg L$*

Finally, we introduce the notion of a dual model. It is obtained from a model (M, T) of **KF** by redefining the extension of the truth predicate. The new extension is defined as the set of all M-sentences, whose negations are not in T (cf. Definition 11).

Definition 15 *For $(M, T) \models KF$, we define:*

- $T^d = Sent - T^-$
- $M^d = (M, T^d)$

Note that T^d may be different from T: in particular, it will contain all sentences which were left indeterminate in the original model (i.e. sentences φ such that neither φ nor $\neg\varphi$ belonged to T).

Useful properties of dual models are described by the theorem below.

Theorem 16.

(a) If $(M, T) \models KF$, then $(M, T^d) \models$ KF1-KF12
(b) If $(M, T) \models KF +$ Cons, then $(M, T^d) \models KF +$ Compl

Proof (chosen cases). Assuming that $(M, T) \models KF +$ Cons, we show:

1 $M^d \models KF13$, i.e. $\forall t \left(T\neg Tt \equiv (\, T(\neg val(t)) \vee \neg \,\mathrm{Sent}_T(val(t)))\right)$.
2 $M^d \models$ Compl.

For 1, we show only ($\leftarrow$). Assume that $M^d \models T\neg t \vee \neg Sent(t)$, so $\neg val(t) \notin T^- \vee \neg Sent(val(t))$; assume also that $M^d \models \neg T\neg Tt$, so $\neg Tt \in T^-$. Then we reason as follows:

(i) $T(t) \in T^+$ (definition of T^-)

[10] Corollary 14 is in fact valid about an *arbitrary* sentence L provably (in $KF +$ Cons or $KF +$ Compl, respectively) equivalent to the statement of its own untruth.

(*ii*) $val(t) \in T^+$ (we know that $(M,T) \models TTt \equiv Tt$)
(*iii*) $\neg val(t) \in T^-$ (definition of T^-)
(*iv*) $\neg Sent(val(t))$ (previous line and our first assumption)
(*v*) $(M,T) \models T(\neg Tt)$ (by $KF13$ and the previous line)
(*vi*) $(M,T) \models \neg T(Tt)$ (by Cons)

So by KF12, $(M,T) \models \neg T(t)$, which means that $val(t) \notin T^+$ - a contradiction.

For 2, we must show: $M^d \models \forall\psi[Sent(\psi) \rightarrow (T(\psi) \vee T(\neg\psi))]$. Fixing a sentence ψ, assume that $\psi \notin T^d$. Then $\psi \in T^-$, so $\neg\psi \in T^+$. By Cons, $\neg\psi \notin T^-$, so $\neg\psi \in T^d$ as required.

4.2 Yablo sentences in KF and some related theories

In this section we investigate properties of formulas which are Yablo in **KF**, in $\mathbf{KF} + \text{Cons}$ and in $\mathbf{KF} + \text{Compl}$ (cf. Definition 1). Our first observation states that the theory $\mathbf{KF} + \text{Compl}$ uniformly decides its Yablo sentences.

Theorem 17. *Let $Y(x)$ be such that $KF + \text{Compl} \vdash Y(x) \equiv \forall z > x \neg T(Y(z))$. Then $KF + \text{Compl} \vdash \forall x \neg Y(x)$.*

Proof. Working in $KF + \text{Compl}$, assume $Y(x)$, so $\forall z > x \neg T(Y(z))$, therefore $\forall z > x+1 \neg T(Y(z))$, so $Y(x+1)$, but also $\neg T(Y(x+1))$. By Fact 13(b), $T(Y(x+1))$ – a contradiction.

When moving to $\mathbf{KF} + \text{Cons}$, things look a bit different. Unlike in the case of $KF + \text{Compl}$ (or **FS**, for that matter) there is no uniform answer to the question "assuming that $Y(x)$ is a Yablo formula in $\mathbf{KF} + \text{Cons}$, does $\mathbf{KF} + \text{Cons}$ prove $Y(n)$?" It turns out that the answers will vary, depending on our choice of $Y(x)$.

Theorem 18. *For every natural number n, there are formulas $Y_0(x)$, $Y_1(x)$ such that:*

(a) Both $Y_0(x)$ and $Y_1(x)$ are Yablo formulas in $\mathbf{KF} + \text{Cons}$.
(b) $\mathbf{KF} + \text{Cons} \vdash Y_0(n)$; $\mathbf{KF} + \text{Cons} \vdash \neg Y_1(n)$

Proof. Let n be fixed; let L be the liar sentence. Define:

- $Y_0(x) := x = n \vee (x > n \wedge L)$
- $Y_1(x) := x = n+1 \vee (x > n+1 \wedge L)$

Then (b) is obviously satisfied. For the proof of (a), we show only that $Y_0(x)$ is a Yablo formula in $\mathbf{KF} + \text{Cons}$ (the argument for $Y_1(x)$ is very similar). Working in $\mathbf{KF} + \text{Cons}$, fix x and consider two cases:
Case 1: $x < n$. Then $\neg Y_0(x)$, and since we also have: $T(n = n \vee (n > n \wedge L)$, we obtain: $\exists z > x T(z = n \vee (z > n \wedge L))$. In effect in Case 1 both sides of the Yablo condition are false, which makes the condition true.
Case 2: $x \geq n$. Since L is provable in $\mathbf{KF} + \text{Cons}$ (Corollary 14), we obtain $Y_0(x)$. And we obtain also the right side of the Yablo condition by the following reasoning: fix $z > x$ and assume $T(Y_0(z))$, i.e. $T(z = n \vee (z > n \wedge L))$. Then by compositional principles of **KF** $T(z = n) \vee (T(z > n) \wedge T(L))$. But by assumption $z > n$; in efect $T(L)$ and therefore $\neg L$ – a contradiction, because L is a theorem of $\mathbf{KF} + \text{Cons}$.

We see in effect, that questions like "does $KF + \textsc{Cons}$ prove $Y(n)$?" do not admit a single answer, independent of our choice of a Yablo formula. In view of this result, narrower classes of Yablo formulas are worth considering. And indeed it turns out that $KF + \textsc{Cons}$ decides a certain narrower, but still quite comprehensive class of Yablo sentences, namely those, which are Yablo in KF itself (without Cons):

Theorem 19. *Let $Y(x)$ be such that $KF \vdash Y(x) \equiv \forall z > x \neg T(Y(z))$. Then $KF + \textsc{Cons} \vdash \forall x Y(x)$.*

Proof. Let $(M,T) \models KF + \textsc{Cons}$. (Then $M^d \models KF + \textsc{Compl}$ – see Theorem 16(b).) For an indirect proof, assume that $(M,T) \models \neg Y(a)$. Fix $b >_M a$ such that $(M,T) \models T(Y(b))$. So $Y(b) \in T^+$; $\neg Y(b) \in T^-$; $\neg Y(b) \notin T^d$. Now we show that:

(*) $\forall z >_M b Y(z) \in T^-$.

Assume that $z >_M b$ and $Y(z) \notin T^-$. So $Y(z) \in T^d$, and (since $z >_M b$) $M^d \models \neg Y(b)$. Therefore (by Fact 13(b) and Theorem 16(b)) $M^d \models T(\neg Y(b))$; in effect $\neg Y(b) \in T^d$, which is a contradiction.
From (*) it follows that $\forall z >_M b Y(z) \notin T^d$, which means that $M^d \models Y(b)$. Therefore $M^d \models Y(b+1)$, so $M^d \models T(Y(b+1))$; but (since $M^d \models Y(b)$) it follows also that $M^d \models \neg T(Y(b+1))$ – a contradiction.

From Theorem 19 it follows directly that if $Y(x)$ is a Yablo formula in **KF**, then $KF + \textsc{Cons} \vdash \forall x > 0 \neg T(Y(x))$. In fact it is possible to show that the formula $Y(0)$ is no exception.

Theorem 20. *If $Y(x)$ is a Yablo formula in* **KF***, then $KF + \textsc{Cons} \vdash \forall z \neg T(Y(z))$.*

Proof. Define $Y^*(x)$ as the formula: $(x = 0 \wedge \forall z \neg T(Y(z))) \vee (x \neq 0 \wedge Y(x-1))$. The theorem is obtained as a direct corollary from Theorem 19 and the fact that $Y^*(x)$ is a Yablo formula in **KF**, i.e. it satisfies provably in **KF**, the usual Yablo condition, i.e.: $Y^*(x) \equiv \forall z > x \neg T(Y^*(z))$.

Given the fact that $Y^*(x)$ is a Yablo formula in **KF**, we can argue as follows. By Theorem 19, $KF + \textsc{Cons} \vdash \forall x Y^*(x)$, so in particular $KF + \textsc{Cons} \vdash \forall x Y^*(0)$, therefore (by the definition of $Y^*(x)$) $KF + \textsc{Cons} \vdash \forall z \neg T(Y(z))$. In effect for the proof of Theorem 20 it is enough to show that we have indeed the Yablo condition for $Y^*(x)$.

For the direction from left to right, assume $Y^*(x)$ and fix $z > x$. Assume $T(Y^*(z))$; then by the definition of $Y^*(x)$ (and by the fact that $z > 0$) we obtain: $T(Y(z-1))$. Then x cannot equal 0, because otherwise by $Y^*(x)$ we would have: $\forall z \neg T(Y(z))$. Since $x \neq 0$, we obtain $Y(x-1)$. But $x - 1 < z - 1$ (because $z > x, x \neq 0$), so $\neg T(Y(z-1))$ - a contradiction.

For the opposite direction, assume $\forall z > x \neg T(Y^*(z))$; assume also $\neg Y^*(x)$, i.e. $\neg[x = 0 \wedge \forall z \neg T(Y(z))] \wedge \neg[x \neq 0 \wedge Y(x-1)]$. Now we consider two cases. *Case 1*: $x = 0$. So $\exists z T(Y(z))$. Fixing such a z and putting $a = z + 1$ we obtain:

$T(a \neq 0 \wedge Y(a-1))$, so $T(Y^*(a))$, which (since $x = 0$) contradicts our main assumption.
Case 2: $x \neq 0$. So $\neg Y(x-1)$, i.e. $\exists z \geq xT(Y(z))$. Fixing such a z and putting $a = z+1$ we obtain again $T(a \neq 0 \wedge Y(a-1))$, i.e. $T(Y^*(a))$, which (since $a > x$) contradicts our main assumption.

From Theorems 19 and 20 it follows easily that every Yablo formula is a liar in $KF + \textsc{Cons}$.

Corollary 21 *If $Y(x)$ is a Yablo formula in* **KF**, *then $KF + \textsc{Cons} \vdash \forall x[Y(x) \equiv \neg T(Y(x))]$.*

Finally, we are going to look at what happens in the theory **KF** itself. The first observation is that it doesn't decide any Yablo sentence:

Corollary 22 *Let $Y(x)$ be a Yablo formula in KF. Then $KF \nvdash \exists x Y(x)$ and $KF \nvdash \exists x \neg Y(x)$.*

Proof. By Theorem 17, the first conjunct is verified by an arbitrary model for $KF + \textsc{Compl}$. By Theorem 19, the second conjunct is verified by an arbitrary model for $KF + \textsc{Cons}$.

In effect each sentence $Y(n)$ is independent of KF.

Does **KF** (without Cons or Compl) settle the issue of equivalence of Yablo sentences? We will show that it does, but for a restricted class of those Yablo sentences, which are well behaved in partial models.

Theorem 23. *Let $Y(x)$ be a Yablo formula in KF such that for every $(M,T) \models KF$ we have (see Definition 11):*

$$\forall a \in M\ [M^* \models_{sk} Y(a) \text{ iff } M^* \models_{sk} \forall z > a \neg T(Y(z))].$$

Then $KF \vdash \forall x Y(x) \vee \forall x \neg Y(x)$.

Proof. Fix a, b such that $(M,T) \models Y(a) \wedge \neg Y(b)$. So we have: $(M,T) \models \forall z > a \neg T(Y(z))$, and also: $(M,T) \models \exists z > bT(Y(z))$.

Let z be the largest number in M such that $(M,T) \models T(Y(z))$. Then $M^* \models_{sk} T(Y(z))$, so (Theorem 12) $M^* \models_{sk} Y(z)$, therefore by the assumptions of the theorem, $M^* \models_{sk} \forall s > z \neg T(Y(s))$. From this we obtain $M^* \models_{sk} \forall s > z+1 \neg T(Y(s))$, and so $M^* \models_{sk} Y(z+1)$ and also $M^* \models_{sk} T(Y(z+1))$. Eventually $(M,T) \models T(Y(z+1)$, which contradicts our choice of z.

From Theorem 23 the following corollary can be easily obtained.

Corollary 24 *For $Y(x)$ satisfying the assumptions of the previous theorem:*

$$KF \vdash \forall xy\ [Y(x) \equiv Y(y)]$$

Finally, we observe that the assumptions of Theorem 23 (and Corollary 24) apply to a class of formulas, which are quite important in the discussions concerning Yablo's paradox (cf. Theorem 2 and its proof).

Observation 25 *Let $Y(x)$ be the formula obtained by diagonalization from the condition $\varphi(x,y) := \forall z > x \neg T(sub(y, name(z))$ (cf. proof of Theorem 2). Then $Y(x)$ satisfies the assumptions of Theorem 23.*

Idea of the proof. As in the standard proof of the diagonal lemma, let $F(x,y)$ be $\varphi(x, subst(y, \ulcorner y \urcorner, name(y)))$; then specify $m = \ulcorner F(x,y) \urcorner$ and define $Y(x)$ as $F(x, \overline{m})$. In effect $Y(x)$ becomes: $\varphi(x, subst(\overline{m}, \ulcorner y \urcorner, name(\overline{m}))$. By performing the substitution operations (interpretation of the truth predicate being irrelevant for the results) it can be verified that $M^* \models_{sk} Y(a)$ iff $M^* \models_{sk} \varphi(a, \ulcorner Y(x) \urcorner)$, which corresponds to the Yablo condition for partial models, as required.

5 Summary

We analysed the behaviour of Yablo formulas in truth theories **FS** and **KF**. It turns out that **FS** proves the equivalence of all Yablo sentences in **FS**. In addition, **FS** treats Yablo formulas as liars: they can be shown to be provably equivalent to the statements of their own untruth.

Theories $KF + \textsc{Cons}$ and $KF + \textsc{Compl}$ both uniformly decide sentences which are Yablo in **KF** (Theorems 17 and 19), although important properties of formulas which are Yablo in $KF + \textsc{Cons}$ depend on the choice of the formula in question (Theorem 18). Yablo formulas obtained by diagonalization in **PAT** are provably equivalent in **KF**.

Acknowledgements. Many thanks to Rafał Urbaniak and Konrad Zdanowski for their useful comments and discussions. The author was supported by a grant from the National Science Centre in Cracow (NCN), decision number DEC-2011/01/B/HS1/03910.

References

1. Beall, J.C.: Is Yablo's Paradox Non-circular? Analysis 61, 176–187 (2001)
2. Cantini, A.: Notes on Formal Theories of Truth. Zeitschrift für Mathematische Logik und Grundlagen der Mathematik 35, 97–130 (1989)
3. Cieśliński, C., Urbaniak, R.: Gödelizing the Yablo Sequence. Journal of Philosophical Logic, published online, DOI 10.1007/s10992-012-9244-4
4. Halbach, V.: Axiomatic Theories of Truth. CUP, Cambridge (2011)
5. Ketland, J.: Yablo's Paradox and ω-Inconsistency. Synthese 145(3), 295–302.
6. Kripke, S.: Outline of a Theory of Truth. Journal of Philosophy 72, 690–716 (1975)
7. Leitgeb, H.: What is a Self-referential Sentence? Critical Remarks on the Alleged (non-)Circularity of Yablo's Paradox. Logique & Analyse 177-178, 3–14 (2002)
8. Priest, G.: Yablo's Paradox. Analysis 57, 236–242 (1997)
9. Sorensen, R.: Yablo's Paradox and Kindred Infinite Liars. Mind 107, 137–155 (1998)
10. Urbaniak, R.: Leitgeb, "about", Yablo. Logique & Analyse 207, 239-254 (2009)
11. Visser, A.: Semantics and the liar paradox. In: Gabbay, D., Guenthner, F. (eds.) Handbook of Philosophical Logic, vol. IV, pp. 617–706. Dordrecht, Kluwer Academic Publishers (1989)
12. Yablo, S.: Paradox Without Self-reference. Analysis 53, 251–252 (1993)