Gödelizing the Yablo Sequence

Cezary Cieśliński · Rafal Urbaniak

Abstract We investigate what happens when 'truth' is replaced with 'provability' in Yablo's paradox. By diagonalization, appropriate sequences of sentences can be constructed. Such sequences contain no sentence decided by the background consistent and sufficiently strong arithmetical theory. If the provability predicate satisfies the derivability conditions, each such sentence is provably equivalent to the consistency statement and to the Gödel sentence. Thus each two such sentences are provably equivalent to each other. The same holds for the arithmetization of the existential Yablo paradox. We also look at a formulation which employs Rosser's provability predicate.

Keywords Incompleteness · Omega-liar · Yablo's paradox · Paradox · Provability · Arithmetic · Goedel

C. Cieśliński
Institute of Philosophy, University of Warsaw, Warsaw, Poland
e-mail: c.cieslinski@uw.edu.pl

R. Urbaniak (✉)
Institute of Philosophy, Sociology and Journalism, Gdansk University, Gdansk, Poland
e-mail: rfl.urbaniak@gmail.com

R. Urbaniak
Centre for Logic and Philosophy of Science, Ghent University, Ghent, Belgium

1 Introduction

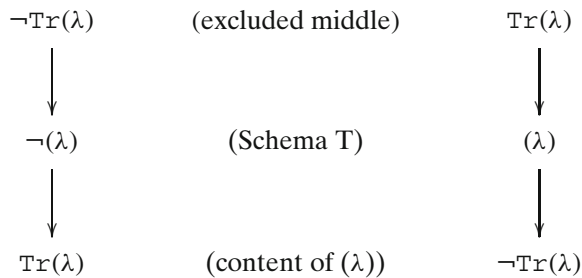1.1 Replacing Truth with Provability

One way to look at the incompleteness of (many) mathematical theories is to compare the Gödelian sentence(s) with those used in various formulations of the liar paradox. The liar sentence says of itself that it is not true:

(λ)   Sentence (λ) is not true.

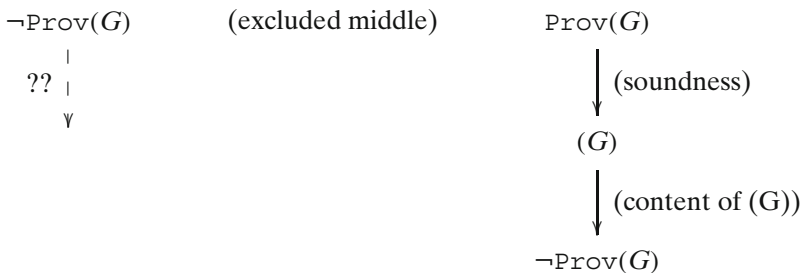A Gödelian sentence (*modulo* syntactic encoding) says of itself that it is not provable:

(G)   Sentence (G) is not provable (in the system).

Thus, Gödelian incompleteness can be seen as a result of turning a paradox into a theorem.[1] Let's sketch and compare the relevant arguments:

$$\neg\text{Tr}(\lambda) \qquad \text{(excluded middle)} \qquad \text{Tr}(\lambda)$$
$$\downarrow \qquad\qquad\qquad\qquad\qquad\qquad \downarrow$$
$$\neg(\lambda) \qquad\qquad \text{(Schema T)} \qquad\qquad (\lambda)$$
$$\downarrow \qquad\qquad\qquad\qquad\qquad\qquad \downarrow$$
$$\text{Tr}(\lambda) \qquad \text{(content of } (\lambda)) \qquad \neg\text{Tr}(\lambda)$$

In the case of the liar, the claim that (λ) is not true turns out to be equivalent to the claim that it is true. Hence the paradox.

What happens when we try to run a parallel argument about provability instead of truth, assuming the soundness of the background theory?[2]

$$\neg\text{Prov}(G) \qquad \text{(excluded middle)} \qquad \text{Prov}(G)$$

$$?? \;\downarrow \qquad\qquad\qquad\qquad\qquad\qquad \downarrow \;\text{(soundness)}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad (G)$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \downarrow \;\text{(content of (G))}$$

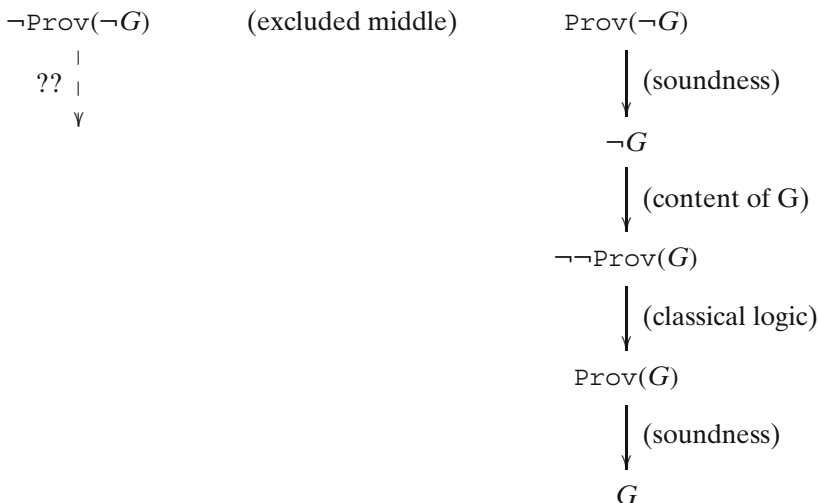$$\qquad\qquad\qquad\qquad\qquad\qquad \neg\text{Prov}(G)$$

---

[1]A fairly common phenomenon. Gödel observed the relationship between his theorem and the liar, and seems to have in fact drawn inspiration from the paradoxes. Other semantic paradoxes have also been used in incompleteness proofs—see e.g. [2], where Grelling's paradox is employed in the proof of the second incompleteness theorem. In ZFC, the proof that there is no universal set can be seen as the result of formalizing Russell's paradox, and both Hartog's theorem and the result that there is no set of all ordinals can been seen as the result of formalizing Burali–Forti.

[2]Of course, the soundness assumption is not needed for the original Gödel's proof to go through, but let's not worry about this.

While the right-hand case goes through, the left-hand case does not get off the ground, because we did not assume that all truths are provable (and thus cannot infer $\neg\phi$ from $\neg\texttt{Prov}(\phi)$). Thus, instead of obtaining a paradox, we just conclude that (G) is not provable in the theory.

The situation is similar with $\texttt{Prov}(\neg G)$:

$\neg\texttt{Prov}(\neg G)$ (excluded middle) $\texttt{Prov}(\neg G)$

$$??\ \Big\downarrow \qquad\qquad\qquad\qquad\qquad\qquad \Big\downarrow \text{(soundness)}$$

$$\neg G$$

$$\Big\downarrow \text{(content of G)}$$

$$\neg\neg\texttt{Prov}(G)$$

$$\Big\downarrow \text{(classical logic)}$$

$$\texttt{Prov}(G)$$

$$\Big\downarrow \text{(soundness)}$$

$$G$$

While from the unprovability of $\neg$G we cannot infer its falsehood, the assumption that it is provable leads to contradiction. Thus, neither G nor $\neg$G are provable in the system, but no paradox ensues.

A question arises as to the extent to which this phenomenon generalizes: are there any other paradoxes involving truth which yield theorems when truth is replaced with provability? We'll take a look at a paradox related to truth which is often claimed to be importantly different from the liar: Yablo's paradox.

1.2 Yablo's Paradox: A Formulation

The paradox has been formulated by Yablo [10]. Consider an infinite sequence of sentences $s_0, s_1, s_2, \ldots$ such that:

$$s_0 = \text{`}\forall x\ (P_1(x) \to \neg\texttt{Tr}(x))\text{'},$$

$$s_1 = \text{`}\forall x\ (P_2(x) \to \neg\texttt{Tr}(x))\text{'},$$

$$s_2 = \text{`}\forall x\ (P_3(x) \to \neg\texttt{Tr}(x))\text{'},$$

$$\vdots$$

$\texttt{Tr}$ is the truth predicate, and the extension of every P_n, (for $n = 1, 2, 3, \ldots$) is $\{s_n, s_{n+1}, s_{n+2}, \ldots\}$. Thus, every s_i says that all s_js with $j > i$ are not true. Now suppose s_0 is true. Then for any $k > 0$, s_k is not true (s_1 is among such formulas, and thus it is not true). Also, s_k is not true for any $k > 1$. But this is exactly what s_1 says—hence s_1 is true after all. Contradiction. Suppose then that s_0 is

false. This means that there is a $k > 0$ such that s_k is true. But we can repeat the reasoning, this time with respect to s_k and reach a contradiction again. No matter whether we assume s_0 to be true or false, we reach a contradiction. Hence the paradox.[3]

2 Yablo Sequence with Provability, a Sketchy Argument

Let's work within some sufficiently strong arithmetical theory **T** and see what happens when `Tr` is replaced with `Prov` in a Yablean sequence. First, we will handwavily describe the kind of reasoning involved, without paying attention to details and without making sure the reasoning can be represented within the system.

Suppose we extend the standard language of arithmetic with a new function symbol '*f*'. We intend the function it expresses to assign to each natural number the g.n. (Gödel number) of the *n*-th sentence in the Yablean sequence (with provability instead of truth):

$$f(n) = \ulcorner \forall x > \bar{n} \; \neg\texttt{Prov}(f(x)) \urcorner \tag{1}$$

(The notation $\bar{n}$ is used for the numeral denoting n.)

Now suppose **T** proves the formula with g.n. $f(n)$:

$$\textbf{T} \vdash \forall x > \bar{n} \; \neg\texttt{Prov}(f(x)) \tag{2}$$

It follows that **T** proves a weakening of Eq. 2:

$$\textbf{T} \vdash \forall x > \overline{n+1} \; \neg\texttt{Prov}(f(x)) \tag{3}$$

and it proves a particular instance of Eq. 2:

$$\textbf{T} \vdash \neg\texttt{Prov}(f(\overline{n+1})) \tag{4}$$

But Eq. 3 simply is the formula with g.n. $f(n+1)$. Also, for any theorem of **T**, **T** being sufficiently strong, proves also that it proves it (see Fact 5 on page 6). So, from Eq. 3, we obtain:

$$\textbf{T} \vdash \texttt{Prov}(f(\overline{n+1}))$$

This, however, together with Eq. 4 shows that **T** is inconsistent. So, if **T** is consistent, Eq. 2 is false and **T** doesn't prove the formula with g.n. $f(n)$ after all.

Let's see what would happen if **T** proved its negation. Suppose

$$\textbf{T} \vdash \neg\forall x > \bar{n} \; \neg\texttt{Prov}(f(x)) \tag{5}$$

---

[3]Let us note that in contrast with the usual liar-type paradoxes based on an explicit use of direct or indirect self-reference, Yablo paradox seems to involve no self-referential loops. The question whether the paradox is really self-reference free has been much debated in recent literature. See e.g. [1, 5, 6, 9].

This yields

$$\mathbf{T} \vdash \exists x > \bar{n} \ \mathtt{Prov}(f(x)) \tag{6}$$

Then there is an m which witnesses this claim:[4]

$$\mathbf{T} \vdash \mathtt{Prov}(f(\bar{m})) \tag{7}$$

But then, $\mathbf{T}$ also proves the formula itself (see e.g. Fact 14 on page 8):

$$\mathbf{T} \vdash \forall x > \bar{m} \ \neg\mathtt{Prov}(f(x)) \tag{8}$$

Now, we can apply to Eq. 8 the same reasoning we applied to Eq. 2 to reach a contradiction. So, both Eqs. 2 and 5 are false and $\mathbf{T}$ doesn't decide the formula with g.n. $f(n)$.

This argument as it stands is not satisfactory. Ideally, we would like to avoid reference to a new function symbol and formulate the reasoning in the original language of arithmetic. We would also like to be clear on what assumptions are being made and on how the steps involved are justified. To this goal, in Section 3 we will first survey a few facts needed for our proof (a reader familiar with the topic can safely skip this section and refer to it later if need be). Then, in Section 4, we will give the full proof.

3 (Slightly) Technical Preliminaries

We assume the reader is familiar with the general framework of Gödelian proofs, but for the sake of accessibility, we list the main results which will be referred to later on. We do it (almost) without proofs, which are rather well-known and should not lie in the focus of our paper. For details we refer the reader to [3] or [7].

Definition 1 An arithmetical theory $\mathbf{T}$ is *p.r. axiomatized* iff all the following conditions hold: (i) the numerical properties of being the g.n. (Gödel numbers) of a $\mathtt{T}$-well formed formula and a $\mathtt{T}$-sentence are p.r. (primitive recursive), (ii) the numerical property of being the g.n. of an axiom is p.r., (iii) the numerical property of being the g.n. of a correct proof is p.r.. An arithmetical theory is *nice* iff it is consistent, p.r. axiomatized, and extends Robinson arithmetic $\mathbf{Q}$.

We will fix our attention on an arbitrary arithmetical theory (in the language of Peano arithmetic) $\mathbf{T}$ supposing it is nice.

---

[4]A full justification of this move hinges on special properties of the formula "$\mathtt{Prov}(x)$" (cf. Fact 3 and Fact 16). It is not true in general that if $\mathbf{T} \vdash \exists x \varphi(x)$, then for some m, $\mathbf{T} \vdash \varphi(m)$.

Fact 2 *Any nice theory* **T** *correctly decides* Δ_0-*sentences.*[5]

The language of **T** can express the property of being a proof in **T**. In particular, there is a Σ_1 formula `Prf(x, y)` which holds of numbers n and m just in case n is a g.n. (Gödel number) of a sequence of formulas which is an **T**-proof of the formula whose g.n. (Gödel number) is m. Moreover, `Prf` captures (binumerates) **T**-provability in **Q**, and therefore also in **T**:

Fact 3 *For any* m *and* n: (1) *If* `Prf`(m, n), *then* **Q** $\vdash$ `Prf`$(\bar{m}, \bar{n})$, (2) *If* $\neg$`Prf`(m, n), *then* **Q** $\vdash \neg$`Prf`$(\bar{m}, \bar{n})$.

Further on, within the language of **T** we can define:

$$\mathtt{Prov}(x) =_{\mathrm{def}} \exists y \; \mathtt{Prf}(y, x)$$

Thus, `Prov` expresses the property of **T**-provability. Clearly, `Prov`(x) is also Σ_1. (Henceforth, we will be using `Prov` sometimes within the language of **T** and sometimes outside of it; sometimes we will also relativize the provability predicate to a given theory. But we trust the context will make it clear what is meant.) Observe the following:

Fact 4 `Prov` *expresses provability:* m *is the g.n. of a* **T***-theorem just in case* `Prov`(m) *is true. Yet, the predicate does not capture provability in* **T***: it is not the case that for any* m*, if* $\neg$`Prov`(m) *then* **T** $\vdash \neg$`Prov`(m).

`Prov`, standardly constructed, has certain properties which will come in handy later on:

Fact 5 (Derivability condition 1)

(D1)　*If* **T** $\vdash \phi$, *then* **T** $\vdash$ `Prov`($\ulcorner\phi\urcorner$). ⊣

Fact 6 (Derivability conditions 2 and 3) *For any nice arithmetical theory* **T** *extending* $\mathbf{I\Sigma_1}$*, there are predicates* `Prf` *and* `Prov` *such that* `Prf` *represents the relation of being a proof in* **T***,* `Prov`(x) *is defined by '*$\exists y \; \mathtt{Prf}(y, x)$*', and* `Prov` *satisfies (provably in* **T***) the following derivability conditions:*

(D2)　**T** $\vdash$ `Prov`($\ulcorner\phi \rightarrow \psi\urcorner$) $\rightarrow$ (`Prov`($\ulcorner\phi\urcorner$) $\rightarrow$ `Prov`($\ulcorner\psi\urcorner$))
(D3)　**T** $\vdash$ `Prov`($\ulcorner\phi\urcorner$) $\rightarrow$ `Prov`($\ulcorner$`Prov`($\ulcorner\phi\urcorner$)$\urcorner$) ⊣

Observe that Fact 6 states only the *existence* of formulas `Prf` and `Prov` with the indicated properties. It does not state that **T** will prove (D2) and

---

[5]For a proof, see e.g. Section 9.7 of [7].

(D3) independently of the choice of the formula `Prf` representing the proof relation. Indeed, such a claim would be false. There are known examples of formulas representing the proof relation which yield provability predicates which do not satisfy (D2) and (D3) (Rosser's provability predicate belongs to this category). However, both (D2) and (D3) hold under the standard construction of the provability predicate.

We use Feferman's dot notation. Within a quotation term, $\dot{z}$ is a term which denotes the numeral for z and depends on the particular choice of z. Also, for any ϕ, we abbreviate $\texttt{Prov}(sub(\ulcorner\phi(x)\urcorner, \dot{z}))$ as $\texttt{Prov}(\phi(z))$ or $\texttt{Prov}(\phi(\dot{z}))$ if we want to emphasize that z is kept free. In fact (a result due to Feferman), for the standard provability predicate, derivability conditions can be slightly generalized:

Fact 7 *Suppose* **T** *is nice and extends* $\mathbf{I\Sigma_1}$. *If* $\mathbf{T} \vdash \phi(x)$, *then* $\mathbf{T} \vdash \texttt{Prov}(\ulcorner\phi(\dot{x})\urcorner)$. *Also,* (D2) *extends to formulas with free variables, and the third condition generalizes to* $\mathbf{T} \vdash \texttt{Prov}(\ulcorner\phi(\dot{x})\urcorner) \to \texttt{Prov}(\ulcorner\texttt{Prov}(\ulcorner\phi(\dot{x})\urcorner)\urcorner)$.

In cases where Fact 7 applies, it validates the move (within a theory) from $\phi \to \psi$ to $\texttt{Prov}(\ulcorner\phi\urcorner) \to \texttt{Prov}(\ulcorner\psi\urcorner)$ and to $\texttt{Prov}(\ulcorner\phi\urcorner) \to \texttt{Prov}(\ulcorner\texttt{Prov}(\ulcorner\psi\urcorner)\urcorner)$ (of course, some attention has to be paid to free variables). We call these moves **K1** and **K2** respectively, and employ them in our proof of Theorem 22.

The consistency statement Con(**T**), expressing consistency of **T**, is defined by $\neg\texttt{Prov}(\ulcorner 0 \neq 0 \urcorner)$, with `Prov` being an appropriate **T**-provability predicate. In what follows we will write often `Con`, suppressing the information about **T**, when the intended theory is clear from the context. We will need a fact which tells us, more or less, that when looking at the consistency statement, it doesn't really make a difference which contradiction we pick.

Fact 8 *If* **T** *is nice and extends* $\mathbf{I\Sigma_1}$, *then for any* ϕ, $\mathbf{T} \vdash [\texttt{Prov}(\phi) \wedge \texttt{Prov}(\neg\phi)] \equiv \neg\texttt{Con}(\mathbf{T})$.

One more fact about `Con` will be useful.

Fact 9 *If* **T** *is nice and derivability conditions hold, then:*

- $\mathbf{T} \vdash \texttt{Con} \equiv \texttt{G}$, *where* `G` *is the Gödel sentence.*
- $\mathbf{T} \vdash \neg\texttt{Prov}(\ulcorner\phi\urcorner) \to \texttt{Con}$ *for any* ϕ.

Definition 10 An arithmetical theory **T** is ω-*consistent* iff for no open formula $\phi(x)$, $\mathbf{T} \vdash \phi(\bar{m})$ for each standard numeral m, and yet $\mathbf{T} \vdash \neg\forall x\, \phi(x)$ (alternatively: iff there is no $\phi(x)$ such that $\mathbf{T} \vdash \exists x\, \phi(x)$ and yet for each m, $\mathbf{T} \vdash \neg\phi(\bar{m})$).

Definition 11 An arithmetical theory **T** is *1-consistent* just in case for no Δ_0 formula $\phi(x)$, $\mathbf{T} \vdash \exists x\, \phi(x)$ and yet for each m, $\mathbf{T} \vdash \neg\phi(\bar{m})$.

The following is a straightforward observation based on Definitions 10 and 11:

Fact 12 *ω-consistency is properly stronger than 1-consistency.*

Definition 13 An arithmetical theory **T** is Σ_1-sound just in case for any Σ_1 sentence, if **T** $\vdash \phi$, then ϕ is true in the standard model of arithmetic.

Fact 14 *If* **T** *is ω-consistent, then if* **T** $\vdash$ `Prov`$(\ulcorner\phi\urcorner)$, *then* **T** $\vdash \phi$.[6]

Fact 15 *If an arithmetical theory* **T** *is nice, then it is 1-consistent iff it is Σ_1-sound.*

Fact 16 *If $\phi(x)$ is a Σ_1-formula and* **T** *is a 1-consistent nice theory, then if* **T** $\vdash$ $\exists x \ \phi(x)$, *then for some m,* **T** $\vdash \phi(\bar{m})$. *(The claim generalizes to formulas with more free variables.)*

Since this fact is not obvious and important in further proofs, here's an argument:

Proof By the niceness of **T** and its 1-consistenty, Fact 15 allows us to infer that **T** is Σ_1-sound. Suppose **T** $\vdash \exists x \ \phi(x)$. If $\phi(x)$ is Σ_1, so is $\exists x \ \phi(x)$. So Σ_1-soundness yields that $\exists x \ \phi(x)$ is true in the standard model. But then, there is a (standard) number m such that $\phi(m)$. We can assume $\phi(m)$ has the form:

$$\exists x_1, \ldots, x_k \ \phi'(m, x_1, \ldots, x_k)$$

where ϕ' is Δ_0. Since it is true in the standard model, there are standard witnesses for the existential quantifiers: some $n_1, \ldots, n_k$ such that $\phi'(m, n_1, \ldots, n_k)$ holds in the standard model. That being the case, Fact 2 allows us to infer:

$$\mathbf{T} \vdash \phi'(\bar{m}, \bar{n}_1, \ldots, \bar{n}_k) \tag{9}$$

By existential generalization we get:

$$\mathbf{T} \vdash \phi(\bar{m}) \tag{10}$$

which completes the proof. (The claim generalizes to ϕ with more free variables.)   □

Fact 16 helps us to strengthen Fact 14 to the following:

Fact 17 *If a nice theory* **T** *is 1-consistent, then if* **T** $\vdash$ `Prov`$(\ulcorner\phi\urcorner)$, *then* **T** $\vdash \phi$.

---

[6]This we have to distinguish from proving instances of reflection schema, `Prov`$(\ulcorner\phi\urcorner) \to \phi$. No nice theory can prove all such instances.

One last thing we need will be a generalized version of the diagonal lemma:

Fact 18 (Diagonal lemma) *If* **T** *is nice, then for any formula* $\phi(x, y)$ *there is a formula* $\psi(x)$ *such that* $\mathbf{T} \vdash [\psi(x) \equiv \phi(x, \ulcorner\psi(x)\urcorner)]$.

Now we are ready to cope with the arithmetized proof.

4 Yablo's Sequence with Provability, Arithmetized

Define ϕ as follows:

$$\phi(x, y) =_{\text{def}} \forall z \ [z > x \rightarrow \neg\texttt{Prov}(sub\,(y, \dot{z}))] \tag{11}$$

with *sub* standing for a substitution function. In effect $\phi(x, y)$ says that for any number z greater than x, the formula whose g.n. is y is not provable of z.

Take a nice theory **T** and use Fact 18 and diagonalize on ϕ. This means there is a formula $Y(x)$ (the provability Yablo formula) such that:[7]

$$\mathbf{T} \vdash Y(x) \equiv \forall z \ [z > x \rightarrow \neg\texttt{Prov}(sub\,(\ulcorner Y(x)\urcorner, \dot{z}))] \tag{12}$$

Theorem 19 *If an arithmetical theory* **T** *is nice, then for any n,* $\mathbf{T} \nvdash Y(\bar{n})$. *Moreover, if* **T** *is also 1-consistent, then for any n,* $\mathbf{T} \nvdash \neg Y(\bar{n})$ *(where* $Y(\bar{n})$ *is constructed following Eqs.* 11 *and* 12 *using* **T**'s *provability predicate*).

Proof For the first part of the theorem, suppose $\mathbf{T} \vdash Y(\bar{n})$ for some n. We'll show this leads to contradiction. By Eq. 12:[8]

$$\mathbf{T} \vdash \forall z \ [z > \bar{n} \rightarrow \neg\texttt{Prov}(Y(z))] \tag{13}$$

Eq. 13 entails two things:

$$\mathbf{T} \vdash \neg\texttt{Prov}(Y(\overline{n+1})) \tag{14}$$

$$\mathbf{T} \vdash \forall z \ [z > \overline{n+1} \rightarrow \neg\texttt{Prov}(Y(z))] \tag{15}$$

But Eq. 15 with Eq. 12 entail:

$$\mathbf{T} \vdash Y(\overline{n+1}) \tag{16}$$

Eq. 16 by the first derivability condition (D1) entails:

$$\mathbf{T} \vdash \texttt{Prov}(\ulcorner Y(\overline{n+1})\urcorner) \tag{17}$$

This, given Eq. 14 implies that **T** is inconsistent. So if **T** is consistent, $\mathbf{T} \nvdash Y(\bar{n})$, for any n. This ends the first half of the proof.

---

[7]This method of constructing a Yablo sequence (but for formulas containing the truth predicate) has been employed by Priest [6] and by Ketland [4]. Also, Priest in footnote 4 of his paper mentions that using provability predicate should lead to a proof of the first incompleteness theorem.

[8]From now on, instead of '$\texttt{Prov}(sub\,(\ulcorner Y(x)\urcorner, \dot{z}))$' we will simply write '$\texttt{Prov}(Y(z))$'.

For the second half, suppose **T** is 1-consistent and $\mathbf{T} \vdash \neg Y(\bar{n})$. We employ Eq. 12 and start as follows:

$$\mathbf{T} \vdash \neg\forall z \; [z > \bar{n} \rightarrow \neg\texttt{Prov}(Y(z))]$$

$$\mathbf{T} \vdash \exists z \; [z > \bar{n} \wedge \texttt{Prov}(Y(z))] \tag{18}$$

Since `Prov` is Σ_1, and $>$ is Δ_0, the formula $\exists z \; [z > \bar{n} \wedge \texttt{Prov}(Y(z))]$ is Σ_1. Thus, Fact 16 applies and there is a number k such that:

$$\mathbf{T} \vdash \bar{k} > \bar{n} \wedge \texttt{Prov}(Y(\bar{k}))$$

$$\mathbf{T} \vdash \texttt{Prov}(Y(\bar{k})) \tag{19}$$

From this, 1-consistency and Fact 17 allow us to infer:

$$\mathbf{T} \vdash Y(\bar{k}) \tag{20}$$

to which we can apply the reasoning used in the first half of the proof, obtaining a contradiction. Thus $\mathbf{T} \nvdash \neg Y(n)$.

(Observe (D2) and (D3) are not used, so the result holds even for some theories without induction and is not sensitive to the construction of the proof predicate.) □

So, when we arithmetize Yablo's paradox using provability instead of truth we get another incompleteness proof, which *prima facie* doesn't use a formula which "says" *of itself* that it is not provable. (Although, if derivability conditions (D1–D3) hold, they are all provably equivalent to such a sentence).

5 Equivalence of Yablo Sentences

First, keep Y relativized to **T**-provability, defined as in Eq. 12,[9] and consider the following claim:

Fact 20 *If* **T** *is nice, then for any* m, n, *if* $m > n$, *then* $\mathbf{T} + Y_T(\bar{n}) \vdash Y_T(\bar{m})$.

The claim is easily provable from the construction of $Y_T(x)$. However, given that in the original paradox, *prima facie* no sentence in the sequence seems to entail a sentence earlier on the list, an obvious question comes to mind:

If **T** is 1-consistent and nice, is it the case that for any m, n, if $m > n$, then $\mathbf{T} + Y_T(\bar{m}) \nvdash Y_T(\bar{n})$?

---

[9]To remind the reader about this, we will sometimes use 'T' in the subscript, but only in contexts where ambiguity may arise.

This would mean that **T**, if 1-consistent and nice, has an infinite sequence of extensions: $\mathbf{T} + Y_T(0)$, $\mathbf{T} + Y_T(1)$, $\mathbf{T} + Y_T(2)$, each weaker than the previous ones. As it turns out, the answer to the question is negative.

Theorem 21 *If* **T** *is nice and* (D1) *and* (D2) *hold, then*

$$\mathbf{T} \vdash \forall x \, [Y(x) \rightarrow \mathtt{Con}(\mathbf{T})]. \;\; \dashv$$

Proof

w.t.s.:	$\mathbf{T} \vdash Y(x) \rightarrow \mathtt{Con}(\mathbf{T})$ (for arbitrary x)
Step 1	(by classical logic) $\mathbf{T} \vdash \bot \rightarrow Y(x+1)$
Step 2	(from Step 1 by (D1)) $\mathbf{T} \vdash \mathtt{Prov}(\ulcorner \bot \rightarrow Y(x+1) \urcorner)$
Step 3	(directly from (D2)) $\mathbf{T} \vdash \mathtt{Prov}(\ulcorner \bot \rightarrow Y(x+1) \urcorner) \rightarrow [\mathtt{Prov}(\ulcorner \bot \urcorner) \rightarrow \mathtt{Prov}(\ulcorner Y(x+1) \urcorner)]$
Step 4	(from Steps 2, 3, by modus ponens) $\mathbf{T} \vdash \mathtt{Prov}(\ulcorner \bot \urcorner) \rightarrow \mathtt{Prov}(\ulcorner Y(x+1) \urcorner)$
Step 5	(from Step 4 by contraposition) $\mathbf{T} \vdash \neg\mathtt{Prov}(\ulcorner Y(x+1) \urcorner) \rightarrow \neg\mathtt{Prov}(\ulcorner \bot \urcorner)$
Step 6	(from Step 5 by definition of $\mathtt{Con}(\mathbf{T})$) $\mathbf{T} \vdash \neg\mathtt{Prov}(\ulcorner Y(x+1) \urcorner) \rightarrow \mathtt{Con}(\mathbf{T})$
Step 7	(by Eq. 12) $\mathbf{T} \vdash Y(x) \rightarrow \neg\mathtt{Prov}(\ulcorner Y(x+1) \urcorner)$
Step 8	(from Steps 6 and 7) $\mathbf{T} \vdash Y(x) \rightarrow \mathtt{Con}(\mathbf{T})$

□

Theorem 22 *For any n, if* **T** *is nice and predicate* $\mathtt{Prov}$ *used in the characterization of the Yablo sequence satisfies the derivability conditions* (D1–D3)*, then* $\mathbf{T} \vdash \forall x \, [\mathtt{Con}(\mathbf{T}) \rightarrow Y(x)]. \; \dashv$

Proof We give a short sketch which uses Gödel's second theorem, and a longer proof which doesn't. For the sketch, working in **T**, assume $\mathtt{Con}(\mathbf{T})$. For an indirect proof, assume also $\neg Y(x)$. This yields $\exists u > x \, \mathtt{Prov}(Y(u))$. Let's fix such an u. Therefore by (D2), $\mathtt{Prov}(\ulcorner \forall y > u \, \neg\mathtt{Prov}(Y(y)) \urcorner)$, so (D2 again) the theory proves that it doesn't prove something. By Fact 9, $\mathtt{Prov}(\ulcorner \mathtt{Con}(\mathbf{T}) \urcorner)$. In effect, by Gödel's second incompleteness theorem formalized in **T**, $\neg\mathtt{Con}(\mathbf{T})$ (the last move requires all three derivability conditions).

Before we give the longer version, the reader should check Fact 7 and our comments about abbreviations and inferential moves which directly precede

and follow our statement of this fact. Also, recall that **K1** depends on the first and second derivability conditions while **K2** depends on all three of them.

w.t.s.:	$\mathbf{T} \vdash \mathrm{Con}(\mathbf{T}) \to Y(x)$ (for arbitrary x)
Step 1	(by Eq. 12) $\mathbf{T} \vdash Y(x) \to \forall z\ [z > x \to \neg\mathrm{Prov}(Y(z))]$
Step 2	(from Step 1 by K2) $\mathbf{T} \vdash \mathrm{Prov}(\ulcorner Y(\dot{x})\urcorner) \to \mathrm{Prov}(\ulcorner \forall z\ [z > \dot{x} \to \neg\mathrm{Prov}(Y(z))]\urcorner)$
Step 3	(Arithmetic) $\mathbf{T} \vdash \forall z\ [z > x \to \neg\mathrm{Prov}(Y(z))] \to \forall z\ [z > x+1 \to \neg\mathrm{Prov}(Y(z))]$
Step 4	(from Step 3 by Eq. 12) $\mathbf{T} \vdash \forall z\ [z > x \to \neg\mathrm{Prov}(Y(z))] \to Y(x+1)$
Step 5	(from Step 4 by K2) $\mathbf{T} \vdash \mathrm{Prov}(\ulcorner \forall z\ [z > \dot{x} \to \neg\mathrm{Prov}(Y(z))]\urcorner) \to$ $\mathrm{Prov}(\ulcorner \mathrm{Prov}(Y(\dot{x}+1))\urcorner)$
Step 6	(Arithmetic) $\mathbf{T} \vdash \forall z\ [z > x \to \neg\mathrm{Prov}(Y(z))] \to \neg\mathrm{Prov}(Y(\dot{x}+1))$
Step 7	(from Step 6 by K1) $\mathbf{T} \vdash \mathrm{Prov}(\ulcorner \forall z\ [z > \dot{x} \to \neg\mathrm{Prov}(Y(z))]\urcorner) \to$ $\mathrm{Prov}(\ulcorner \neg\mathrm{Prov}(Y(\dot{x}+1))\urcorner)$
Step 8	(Logic, Steps 5 and 7) $\mathbf{T} \vdash \mathrm{Prov}(\ulcorner \forall z\ [z > \dot{x} \to \neg\mathrm{Prov}(Y(z))]\urcorner) \to$ $\mathrm{Prov}(\ulcorner \mathrm{Prov}(Y(\dot{x}+1))\urcorner) \wedge \mathrm{Prov}(\ulcorner \neg\mathrm{Prov}(Y(\dot{x}+1))\urcorner)$
Step 9	(from Step 8 by Fact 8) $\mathbf{T} \vdash \mathrm{Prov}(\ulcorner \forall z\ [z > \dot{x} \to \neg\mathrm{Prov}(Y(z))]\urcorner) \to \neg\mathrm{Con}(\mathbf{T})$
Step 10	(Logic, Steps 2 and 9) $\mathbf{T} \vdash \mathrm{Con}(\mathbf{T}) \to \neg\mathrm{Prov}(\ulcorner Y(\dot{x})\urcorner)$
Step 11	(Logic, Step 10) $\mathbf{T} \vdash \mathrm{Con}(\mathbf{T}) \to (\forall z)\neg\mathrm{Prov}(\ulcorner Y(z)\urcorner)$
Step 12	(Logic, Steps 11 and Eq. 12) $\mathbf{T} \vdash \mathrm{Con}(\mathbf{T}) \to Y(x)$

$\square$

From Theorems 21 and 22 the following corollaries follow:

Corollary 23 *If* **T** *is nice and the derivability conditions are satisfied, then* $\mathbf{T} \vdash \forall x, y\ [Y(x) \equiv Y(y)]$.

Corollary 24 *The answer to our question is pretty much negative. As long as* **T** *is nice and the derivability conditions are satisfied,* $\mathbf{T} + Y_T(\bar{m}) \vdash Y_T(\bar{n})$, *and it doesn't matter whether* $m > n$.

Also, recall that according to Fact 9, the Gödel sentence is provably equivalent to the consistency statement. Thus:

Corollary 25 *If* **T** *is nice and the derivability conditions hold,* **T** $\vdash$ G $\equiv Y(\bar{n})$ *for any n.*

6 Existential Yablo Paradox

It seems we can run a similar paradox, but with existential quantification (see [8]). Consider the infinite sequence:

$$s_0 = \text{`}\exists x \, (P_1(x) \wedge \neg \text{Tr}(x))\text{'},$$
$$s_1 = \text{`}\exists x \, (P_2(x) \wedge \neg \text{Tr}(x))\text{'},$$
$$s_2 = \text{`}\exists x \, (P_3(x) \wedge \neg \text{Tr}(x))\text{'},$$
$$\vdots$$

(where the reference of P_is is as in the original formulation). Suppose s_0 is true. Then, a sentence above it in the list, say s_k, has to be false. But then, it is not the case that $\exists x > k \, \neg \text{Tr}(s_x)$. That is, $\forall x > k \, \text{Tr}(s_x)$. In particular, $\text{Tr}(s_{k+1})$ which says that there is a false sentence above s_{k+1}. But it's also above s_k, and all sentences above s_k are to be true. Contradiction. So s_0 is not true. Suppose it is false. Then, we just run for s_0 the second half of the above reasoning.

What happens when we arithmetize this paradox replacing Tr with Prov? Say we introduce

$$\psi(x, y) =_{\text{def}} \exists z > x \, \neg \text{Prov}(sub(y, \dot{z})) \tag{21}$$

Once we diagonalize on ψ we know that there is a formula $E(x)$ such that:

$$\textbf{T} \vdash E(x) \equiv \exists z > x \, \neg \text{Prov}(sub(\ulcorner E(x) \urcorner, \dot{z})) \tag{22}$$

Theorem 26 *If* **T** *is nice and 1-consistent, then* **T** $\nvdash \neg E(\bar{n})$, *for any standard n.*

Proof Suppose **T** $\vdash \neg E(\bar{n})$. Then:[10]

$$\textbf{T} \vdash \forall z > \bar{n} \, \text{Prov}(E(z)) \tag{23}$$

$$\textbf{T} \vdash \text{Prov}(\ulcorner E(\overline{n+1}) \urcorner)$$

By Fact 17 and 1-consistency of **T**:

$$\textbf{T} \vdash E(\overline{n+1})$$

---

[10]'$\text{Prov}(E(z))$' abbreviates '$\text{Prov}(sub(\ulcorner E(x) \urcorner, \dot{z}))$'.

This means:

$$\mathbf{T} \vdash \exists z > \overline{n+1} \ \neg\mathtt{Prov}(E(z))$$

But Eq. 23 also entails:

$$\mathbf{T} \vdash \forall z > \overline{n+1} \ \mathtt{Prov}(E(z))$$

which shows **T** to be contradictory. □

Further, say we assume for *reductio* that $\mathbf{T} \vdash E(\bar{n})$. We obtain:

$$\mathbf{T} \vdash \exists z > \bar{n} \ \neg\mathtt{Prov}(\ulcorner E(z) \urcorner) \tag{24}$$

This doesn't allow us to infer that there is a standard witness for which this claim is provable. We can try to continue reasoning within **T** with an arbitrary constant (just making sure we don't apply any principles meant to apply only to standard numerals):

$$a > \bar{n} \wedge \neg\mathtt{Prov}(\ulcorner E(a) \urcorner)$$

But here the argument is blocked. Contrary to what happened with the original proof, Eq. 24 does not entail the content of $E(a)$. Yet, there is another, rather simple argument and the claim holds.

Theorem 27 *If* **T** *is nice,* $\mathbf{T} \nvdash E(\bar{n})$*, for any n.*

Proof If $\mathbf{T} \vdash E(\bar{n})$, it proves that it doesn't prove something, and thus proves its own consistency (by Fact 9). But if it is nice, it doesn't do the latter. □

Now we can ask whether all E-sentences are provably equivalent the way all Y-sentences are. The answer is positive:

Theorem 28 *If* **T** *is nice and the derivability conditions are satisfied, all E-sentences are provably equivalent in* **T**. *That is, for any n and m:*

$$\mathbf{T} \vdash E(\bar{n}) \equiv E(\bar{m}). \ \dashv$$

Proof The proof is by provable equivalence of each such sentence with $\mathtt{Con}(\mathbf{T})$. From left to right, $E(\bar{n}) \to \mathtt{Con}(T)$, this is clear because $E(\bar{n})$ states the unprovability of a sentence, and this implies the consistency statement (Fact 9).

From right to left, work within **T**. Assume $\mathtt{Con}(\mathbf{T})$ and $\neg E(\bar{n})$. This gives $\forall z > n \ \mathtt{Prov}(\ulcorner E(z) \urcorner)$. Thus $\mathtt{Prov}(\ulcorner E(\overline{n+1}) \urcorner)$. So we get $\mathtt{Prov}(\ulcorner \mathtt{Con}(\mathbf{T}) \urcorner)$, because we already have the implication $E(k) \to \mathtt{Con}(\mathbf{T})$ within the scope of $\mathtt{Prov}$ for any k. But Gödel's second theorem (formalized) gives $\mathtt{Con}(\mathbf{T}) \to \neg\mathtt{Prov}(\ulcorner \mathtt{Con}(\mathbf{T}) \urcorner)$. So $\neg\mathtt{Prov}(\ulcorner \mathtt{Con}(T) \urcorner)$, which gives a contradiction and ends the proof. □

The way the proof proceeds makes the following clear:

Corollary 29 *All E-sentences are provably equivalent to the consistency statement and to the Gödel sentence, if the background theory is nice and the derivability conditions are satisfied.*

7 Yablo Formulas with Rosser's Provability Predicate

A case deserving special attention is that of Rosser's provability predicate. We introduce two definitions:[11]

$$\texttt{Prf}^{\text{R}}(x, y) =_{\text{def}} \texttt{Prf}(x, y) \wedge \forall z < x \neg \texttt{Prf}(z, \neg y), \tag{25}$$

$$\texttt{Prov}^{\text{R}}(y) =_{\text{def}} \exists x \texttt{Prf}^{\text{R}}(x, y). \tag{26}$$

Then it is possible to construct a Yablo formula $Y^R(x)$ satisfying:

$$\mathbf{T} \vdash Y^R(x) \equiv \forall z > x \neg \texttt{Prov}^{\text{R}}(Y^R(z))]. \tag{27}$$

Since $\texttt{Prov}^{\text{R}}(y)$ satisfies (D1), Theorem 19 applies: our theory $\mathbf{T}$ doesn't decide $Y^R(n)$ for any natural number n. However, $\texttt{Prov}^{\text{R}}(y)$ doesn't satisfy (D2) and (D3)[12] and so Corollaries 23 and 24 do not apply. The issue whether all $Y^R(n)$-s are provably equivalent remains undecided.

Is it possible to construct a model M for $\mathbf{T}$ with $M \models Y^R(n + 1)$ and $M \nvDash Y^R(n)$? We don't know the answer, but the following is worth observing:

Fact 30 *Any model M of* $\mathbf{T}$ *with* $M \models Y^R(x + 1)$ *and* $M \nvDash Y^R(x)$ *would have to satisfy* $\neg Con(PA)$.

Proof Fix $x \in M$ and suppose:

$$M \models Y^R(x + 1)$$
$$M \models \neg Y^R(x)$$

These (respectively) yield (in the model):

$$\forall z > x + 1 \ \neg \texttt{Prov}^R(Y^R(z)) \tag{28}$$

$$\exists z > x \ \texttt{Prov}^R(Y^R(z)) \tag{29}$$

Together, Eqs. 28 and 29 entail:

$$\texttt{Prov}^R(Y^R(x + 1))$$

---

[11]Of course, '$\neg y$' stands for the g.n. of the negation of the formula whose g.n. is y.

[12]This is the case because (D1)–(D3) are enough for the proof of the second incompleteness theorem. Otherwise we would have: $\mathbf{T} \nvdash \texttt{Con}^{\text{R}}(\mathbf{T})$, with $\texttt{Con}^{\text{R}}(\mathbf{T})$ defined by $\neg \texttt{Prov}^{\text{R}}(0 \neq 0)$. However, as it is well-known, $\mathbf{T} \vdash \texttt{Con}^{\text{R}}(\mathbf{T})$.

By definition of $\texttt{Prov}^R$, it follows:

$$\texttt{Prov}(Y^R(x+1))$$

Now, work inside the scope of the provability predicate. $Y^R(x+1)$ by definition is equivalent to:

$$\forall z > x + 1 \ \neg\texttt{Prov}^R(Y^R(z)) \tag{30}$$

Eq. 30 entails $\forall z > x + 2 \ \neg\texttt{Prov}^R(Y^R(z))$ and by the same token $Y^R(x+2)$. Thus, jumping out of the scope of the provability predicate:

$$\texttt{Prov}(Y^R(x+2)). \tag{31}$$

On the other hand, we already have Eq. 28, which entails:

$$\neg\texttt{Prov}^R(Y^R(x+2))$$

By the construction of Rosser predicate, this means:

$$\forall z \ [\neg\texttt{Prf}(z, \ulcorner Y^R(x+2)\urcorner) \vee \exists s < z \ \texttt{Prf}(s, \ulcorner\neg Y^R(x+2)\urcorner)] \tag{32}$$

We know by Eq. 31 that there is a number d such that $\texttt{Prf}(d, Y^R(x+2))$. Applying Eq. 32 to d, we see that d must satisfy the second disjunct, i.e. $\exists s < d \ \texttt{Prf}(s, \neg Y^R(x+2))$. In effect $\texttt{Prov}(\neg Y^R(x+2))$, which in view of Eq. 31 gives us $\neg\texttt{Con}(\mathbf{T})$. □

8 Summary

Replacing truth with provability in Yablo's paradox yields an infinite sequence of undecided formulas, as long as the background arithmetical theory satisfies standard requirements for incompleteness. All such formulas are provably equivalent to the consistency claim and to the Gödel sentence. These properties are preserved when the original paradox is replaced by the existential Yablo paradox.

When Rosser's provability predicate is used in the construction, independence claims hold but it remains undecided whether in such a case all sentences in a Yablean sequence (i.e. all sentences obtained from '$Y^R(x)$' by substituting numerals for 'x') are provably equivalent in $\mathbf{T}$. We conjecture that such sentences are not provably equivalent, even though they are all true in the standard model of arithmetic.

References

1. Beall, J.C. (2001). Is Yablo's paradox non-circular? *Analysis, 61*, 176–187.
2. Cieslinski, C. (2002). Heterologicality and incompleteness. *Mathematical Logic Quarterly, 48*, 105–110.
3. Hájek, P., & Pudlák, P. (1998). *Metamathematics of first-order arithmetic*. Springer.
4. Ketland, J. (2005). Yablo's paradox and ω-inconsistency. *Synthese, 145*(3), 295–302.
5. 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.
6. Priest, G. (1997). Yablo's paradox. *Analysis, 57*, 236–242.
7. Smith, P. (2007). *An introduction to Gödel's theorem's*. Cambridge University Press.
8. Sorensen, R. (1998). Yablo's paradox and kindred infinite liars. *Mind, 107*, 137–154.
9. Urbaniak, R. (2009). Leitgeb, "about", Yablo. *Logique & Analyse, 207*, 239–254.
10. Yablo, S. (1993). Paradox without self–reference. *Analysis, 53*, 251–252.