Disentangling Truth from Undecidability: the gödelian view

Mario Piazza (m.piazza@unich.it) and Gabriele Pulcini
(gab.pulcini@gmail.com)
Department of Philosophy
University of Chieti-Pescara

Abstract. In many readings of Gödel's first incompleteness theorem it is usual to find the claim that the theorem proves the existence of arithmetic statements which are *true* but unprovable. This intuitive reading has often obscured the understanding of the theorem. In this paper we will argue in favour of a disentanglement of truth from undecidability.

Keywords: Peano Arithmetic, Gödel's incompleteness, mathematical truth

1. Introduction

In the first paragraph of the famous 1931 article, Gödel does expound the first incompleteness theorem by means of an informal explanation [3]. The argument can be rebuilt as follows. Let PA denote the formal system of first order arithmetic and let

$$\varphi \equiv \varphi \textit{ is not provable in } \texttt{PA}$$

be the well-known gödelian proposition; let's now ask ourselves whether φ is provable or not in PA.

- Suppose that φ is provable. Then, for what it literally says, it would be a false statement. This would mean that PA is unsound inasmuch it allows a false statement to be proved. Hence: if PA is sound, then φ is unprovable in it.

- Now suppose that φ is unprovable in PA. Then, it is a true statement and its formal negation $\neg\varphi$ will be false, because the negation of a statement is false if and only if this statement is true and vice versa. Again, if PA is sound, then $\neg\varphi$ is unprovable.

Therefore, there exists a proposition φ such that neither φ nor its negation $\neg\varphi$ are provable in PA, so that PA is *syntactically incomplete*. Moreover, since φ is a true statement, PA is also *semantically incomplete*, i.e. there exists a true statement that PA cannot prove.

This informal argument that Gödel launches as a sort of guide for the perplexed is clearly of a semantical kind to the extent to induce many readers to shift toward semantical aspects instead of strictly dwelling on the syntactical properties of PA in which Gödel frames the scenario of incompleteness in the sequel of the article. Such a change of focus has usually misguided the studies on the incompleteness as to generate and support the myth that the first theorem proves the existence of arithmetical statements which are *true but unprovable* and, therefore, that truth the undecidability in the theorem are entangled, although via a crushing *mismatch*. This popular reading, albeit intuitive, obscures the correct interpretation of the Gödel theorem whereby – as we shall argue in this paper – the disentanglement of truth from undecidability is the very condition of making sense of the incompleteness.

Now, the supposed mismatch between truth and undecidability is subsumed by the endorsement of an interplay between an *outside* dimension of truth (prior to any well-specified deductive argument) and a *inside* dimension of provability (related with a specific deductive system). So the outside conception of truth reflects the philosophical view according to which a certain (mathematical) fact is true or false independently from any possible (constructive or not) verification process arranged into a mathematical proof; hence its true pervades the whole domain of mathematical disciplines without exceptions. Nevertheless, in this paper we show that as the line of reasoning that bears on incompleteness becomes more and more convincing and precise, the corresponding conception of truth ends by losing traces of an outside character. Moreover, we bring this inside view of truth into alignment with Gödel's earlier works on the incompleteness of arithmetic. In particular, we take into account the Princeton lectures of 1934 [4] by focusing on:

- the footnote 21, designed to clear up that the representability of the primitive recursive functions in PA is not strictly required for obtaining undecidable propositions,
- the paragraph 7, entitled *Relations of the foregoing arguments to paradoxes.*

On the one hand, the footnote 21 considers an inside version of truth that essentially corresponds to the provability in elementary number theory; on the other hand, in paragraph 7, Gödel recalls that the disentanglement of the truth notion from that one of provability enables to avoid the riproposition of the liar paradox in his logical construction.

2. Technical Preliminaries

In this section, we recall some basic logical relations holding between the familiar context of elementary number theory – represented by the standard model $\mathcal{N}$ – and its formal counterpart given by PA [6, 7]. These two contexts will be clearly distinguished by writing strings of symbols from PA in typewrite style. The proofs for the incompleteness theorems are here just sketched; the reader can find all the complementary technical details in [9].

DEFINITION 1 (Peano Arithmetic). *The language of* PA *is formed by the usual language of first order logic with identity enriched by the individual constant "*0*", the unary functional symbol "x'" (the successor) and the two binary functional symbols "+" and "·". Moreover,* PA *is deductively defined by the following nine axioms [6].*

1. $\vdash_{\mathtt{PA}} \mathtt{x} = \mathtt{y} \rightarrow (\mathtt{x} = \mathtt{z} \rightarrow \mathtt{y} = \mathtt{z})$
2. $\vdash_{\mathtt{PA}} \mathtt{x} = \mathtt{y} \rightarrow \mathtt{x}' = \mathtt{y}'$
3. $\vdash_{\mathtt{PA}} \mathtt{0} \neq \mathtt{x}'$
4. $\vdash_{\mathtt{PA}} \mathtt{x}' = \mathtt{y}' \rightarrow \mathtt{x} = \mathtt{y}$
5. $\vdash_{\mathtt{PA}} \mathtt{x} + \mathtt{0} = \mathtt{x}$
6. $\vdash_{\mathtt{PA}} \mathtt{x} + \mathtt{y}' = (\mathtt{x} + \mathtt{y})'$
7. $\vdash_{\mathtt{PA}} \mathtt{x} \cdot \mathtt{0} = \mathtt{0}$
8. $\vdash_{\mathtt{PA}} \mathtt{x} \cdot \mathtt{y}' = (\mathtt{x} \cdot \mathtt{y}) + \mathtt{x}$
9. *For every formula $\alpha(x)$ of* PA *such that x occurs free in α,*
   $\vdash_{\mathtt{PA}} \alpha(\mathtt{0}) \rightarrow (\forall \mathtt{x}(\alpha(\mathtt{x}) \rightarrow \alpha(\mathtt{x}')) \rightarrow \forall \mathtt{x} \alpha(\mathtt{x}))$.

DEFINITION 2 (structure $\mathcal{N}$). *The structure $\mathcal{N} = (\mathbb{N}, 0, +, \cdot)$ is formed by the set of non-negative integers $\mathbb{N} = \{0, 1, 2, \ldots\}$, the distinguished number $0 \in \mathbb{N}$ and the usual symbols $+$ and $-$ for sum and product.*

Notation. We abbreviate with $\mathtt{n}$ the numeral $\mathtt{0}''^{\cdots\prime}$ resulting from n applications of the successor function to the constant 0. For any pair of terms t and s, $t \neq s$ is intended to be equivalent to $\neg(t = s)$.

THEOREM 1. *$\mathcal{N} \models$ PA, i.e. $\mathcal{N}$ is a model for* PA.

Proof. The structure $\mathcal{N}$ does interpret PA as follows:

– $\mathtt{0}^{\mathcal{N}} = 0$,

- for each $n \in \mathbb{N}$: $\mathtt{n}^{\mathcal{N}} = n$,
- sum "+" and product "·" in $\mathcal{N}$ respectively interprets their formal counterparts "+" and "·" in PA.

The proof consists in showing that PA is sound w.r.t. $\mathcal{N}$, in other words: if $\vdash_{\mathtt{PA}} \alpha$, then $\mathcal{N} \models \alpha$. We proceed as usual by induction. It is immediate to check that $\mathcal{N}$ satisfies axioms 1-8. As far as the induction principle is concerned (axiom 9), since the domain of $\mathcal{N}$ coincides with the set of naturals $\mathbb{N}$, the inductive mechanism is indeed able to cover the totality of the elements of $\mathcal{N}$ so as to justify the introduction of the universal quantifier.

THEOREM 2. PA *admits* non-standard models*, namely models non-isomorphic to the standard one.*

Proof. Consider the theory $\mathtt{PA}_\infty$ obtained from PA by expanding its language with the constant c and by adding the following infinite set of axioms:

$$\text{for all } n \in \mathbb{N}:\ \vdash_{\mathtt{PA}} \mathtt{c} > \mathtt{n}.$$

Consider now the generic subsystem $\mathtt{PA}_{\infty\downarrow k}$ of $\mathtt{PA}_\infty$ obtained by taking a finite subset of the new axioms as follows:

$$\vdash_{\mathtt{PA}} \mathtt{c} > \mathtt{0}, \quad \vdash_{\mathtt{PA}} \mathtt{c} > \mathtt{1}, \quad , \quad \dots \quad , \quad \vdash_{\mathtt{PA}} \mathtt{c} > \mathtt{k}.$$

The model $\mathcal{N}^* = (\mathbb{N}, 0, k, +, \cdot)$, where $k \in \mathbb{N}$ interprets the constant c, satisfies $\mathtt{PA}_{\infty\downarrow k}$.

Finally, the compactness theorem ensures the existence of a model $\mathcal{M} \models \mathtt{PA}_\infty$ and so $\mathcal{M} \models \mathtt{PA}$. Moreover, $\mathcal{M}$ is clearly non isomorphic to $\mathcal{N}$ since there is an element $a \in |\mathcal{M}| \backslash \mathbb{N}$ "bigger" than any positive integer.

REMARK 1. *Since PA does admit non-isomorphic structures among the range of its models, it is said to be a* non-categorical *theory. It means that the formal system of first order arithmetic is not able to univocally grasp the mathematical structure of* $\mathbb{N}$ *but among a series of other "bizarre" non-standard structures.*

The reader should also remark that, whereas the proof of $\mathcal{N} \models \mathtt{PA}$ *is developed by following the inner deductive mechanism of* PA*,* $\mathcal{M} \models \mathtt{PA}$ *comes through the external contribution of the compactness theorem. This is due to the fact that a recursive argument working axiom by axiom would fail when considering the induction principle expressed by axiom 9. In other words, the inductive mechanism is not able to cover the non-standard elements belonging to* $|\mathcal{M}|$.

PROPOSITION 1. *Let* $n, m \in \mathbb{N}$ *and* t, s, r *closed terms:*

1. *if* $\mathcal{N} \vDash n = m$, *then* $\vdash_{\texttt{PA}} \texttt{n} = \texttt{m}$,

2. *if* $\mathcal{N} \vDash n \neq m$, *then* $\vdash_{\texttt{PA}} \texttt{n} \neq \texttt{m}$,

3. $\vdash_{\texttt{PA}} \texttt{t} = \texttt{s} \rightarrow \texttt{s} = \texttt{t}$,

4. $\vdash_{\texttt{PA}} \texttt{t} = \texttt{r} \rightarrow (\texttt{t} = \texttt{s} \rightarrow \texttt{r} = \texttt{s})$.

Proof. The reader can find all the proofs in [6].

LEMMA 1. *Let t, s be two closed terms:*

1. *if* $\mathcal{N} \vDash t = s$, *then* $\vdash_{\texttt{PA}} \texttt{t} = \texttt{s}$,

2. *if* $\mathcal{N} \vDash t \neq s$, *then* $\vdash_{\texttt{PA}} \texttt{t} \neq \texttt{s}$.

Proof. At first we prove a weaker version of the two claims in which $s = n \in \mathbb{N}$.

Let $\mathcal{N} \vDash t = n$, then $\vdash_{\texttt{PA}} \texttt{t} = \texttt{n}$ can be easily proved by induction on the number of sums and products occurring in t and Proposition 1.1 gives the basis. Now suppose that, for a certain pair $n, m \in \mathbb{N}$, $\mathcal{N} \vDash t \neq n$, $\mathcal{N} \vDash t = m$ and $\mathcal{N} \vDash n \neq m$. By the just proved property and Proposition 1.2, it is $\vdash_{\texttt{PA}} \texttt{t} = \texttt{m}$ and $\vdash_{\texttt{PA}} \texttt{n} \neq \texttt{m}$. Then, through Proposition 1.4, we obtain $\vdash_{\texttt{PA}} (\texttt{t} = \texttt{m} \wedge \texttt{n} \neq \texttt{m}) \rightarrow \texttt{t} \neq \texttt{n}$ and so $\vdash_{\texttt{PA}} \texttt{t} \neq \texttt{n}$.

At this point, we generalise these properties as follows.

(Prop. 1.1) Let $\mathcal{N} \vDash t = s$, it means that there is an $n \in \mathbb{N}$ such that $\mathcal{N} \vDash t = n$ and $\mathcal{N} \vDash s = n$. By the just proved property, we deduce both $\vdash_{\texttt{PA}} \texttt{t} = \texttt{n}$ and $\vdash_{\texttt{PA}} \texttt{s} = \texttt{n}$; by Proposition 1.3, it is $\vdash_{\texttt{PA}} \texttt{n} = \texttt{t}$ and $\vdash_{\texttt{PA}} \texttt{n} = \texttt{s}$. We apply Proposition 1.4 obtaining $\vdash_{\texttt{PA}} \texttt{n} = \texttt{t} \rightarrow (\texttt{n} = \texttt{s} \rightarrow \texttt{t} = \texttt{s})$ and consequently $\vdash_{\texttt{PA}} \texttt{t} = \texttt{s}$.

(Prop. 1.2) Let $\mathcal{N} \vDash t \neq s$, it means that there is an $m \in \mathbb{N}$ such that $\mathcal{N} \vDash t = m$ and $\mathcal{N} \vDash s \neq m$; by the just proved properties, we have $\vdash_{\texttt{PA}} \texttt{t} = \texttt{m}$ and $\vdash_{\texttt{PA}} \texttt{s} \neq \texttt{m}$. Through Proposition 1.4, we obtain $\vdash_{\texttt{PA}} (\texttt{t} = \texttt{m} \wedge \texttt{s} \neq \texttt{m}) \rightarrow \texttt{t} \neq \texttt{s}$ and finally $\vdash_{\texttt{PA}} \texttt{t} \neq \texttt{s}$.

DEFINITION 3 (arithmetical hierarchy). *A formula is called Δ_0 if all its quantifiers are bounded. We call a formula Σ_1 (resp. Π_1) if it has the form $\exists x\alpha$ (resp. $\forall x\alpha$) with $\alpha \in \Delta_0$.*

REMARK 2. *Whereas $\alpha \in \Sigma_1$ if, and only if, $\neg\alpha \in \Pi_1$ – the set of Δ_0 formulas is closed under negation. Moreover, it is $\Delta_0 \subset \Sigma_1, \Pi_1$.*

THEOREM 3 (Δ_0-syntactical completeness). PA *is Δ_0-syntactically complete i.e., if $\alpha \in \Delta_0$, then either $\vdash_{\texttt{PA}} \alpha$ or $\vdash_{\texttt{PA}} \neg\alpha$.*

Proof. Let $\alpha \in \Delta_0$; we proceed by induction on the number of logical connectives occurring in α.

Base. If no logical connective does occur in α, then $\alpha \equiv t = s$ with t, s closed terms. It is either $\mathcal{N} \models t = s$ or $\mathcal{N} \models t \neq s$ and so Lemma 1 gives us the basis.

Step. By means of the following conversions

$$\exists x \leq k\alpha(x) \leftrightarrow \alpha(0) \wedge \alpha(1) \wedge \ldots \wedge \alpha(k)$$
$$\forall x \leq k\alpha(x) \leftrightarrow \alpha(0) \vee \alpha(1) \vee \ldots \vee \alpha(k),$$

any quantified Δ_0-formula can be rewritten into an equivalent one without quantifiers. Then it is easy to see that any boolean composition of decidable proposition is, in turn, decidable.

DEFINITION 4 (deductive independence). *A formula α is said to be independent from* PA *if* $\nvdash_{\mathtt{PA}} \alpha$ *and* $\nvdash_{\mathtt{PA}} \neg\alpha$.

DEFINITION 5 (ω-consistency). *A theory* T *is ω-consistent if the following two conditions are mutually excluding:*

- *for all* $n \in \mathbb{N}$, $\vdash_{\mathtt{T}} \alpha(\mathtt{n})$,
- $\vdash_{\mathtt{T}} \exists \mathtt{x} \neg\alpha(\mathtt{x})$.

THEOREM 4 (first incompleteness theorem). *There exists a formula* $\varphi \in \Pi_1$ *such that, if* PA *is ω-consistent, then φ is independent from* PA.

Proof. The proof is developed through the following five points.

1. There exists a 1-1 assignment of natural numbers to formulas and demonstrations of PA. $\ulcorner\alpha\urcorner$ and $\overline{\ulcorner\alpha\urcorner}$ respectively indicate the number associated with α (its gödelian code) and its corresponding numeral.
2. It is possible to define in the language of PA a binary Δ_0-predicate $Dem(x, y)$ such that $\vdash_{\mathtt{PA}} \mathtt{Dem}(\mathtt{n}, \mathtt{m})$ if, and only if, n encodes a demonstration of the formula α with $\ulcorner\alpha\urcorner = m$;
3. Consider the Σ_1-predicate $Theor(y) \equiv \exists x Dem(x, y)$, its negation admits a formula φ as fixed point, i.e.:
$$\vdash_{\mathtt{PA}} \varphi \leftrightarrow \neg\mathtt{Theor}(\overline{\ulcorner\varphi\urcorner}).$$
4. If PA is consistent, then $\vdash_{\mathtt{PA}} \varphi$ implies $\vdash_{\mathtt{PA}} \neg\varphi$ and so $\nvdash_{\mathtt{PA}} \varphi$.
5. If PA is ω-consistent, then $\nvdash_{\mathtt{PA}} \varphi$ implies $\nvdash_{\mathtt{PA}} \neg\varphi$.

THEOREM 5 (second incompleteness theorem). *Consider the formula*

$$Cons_{\texttt{PA}} \equiv \neg Theor(0=1)$$

asserting the consistency of PA*: it is independent from* PA*.*

Proof. The proof consists in showing that $Cons_{\texttt{PA}}$ is provably equivalent to φ, i.e. $\vdash_{\texttt{PA}} \texttt{Cons}_{\texttt{PA}} \leftrightarrow \varphi$. In such a way, $\vdash_{\texttt{PA}} \texttt{Cons}_{\texttt{PA}}$ and $\vdash_{\texttt{PA}} \neg\texttt{Cons}_{\texttt{PA}}$ would respectively imply $\vdash_{\texttt{PA}} \varphi$ and $\vdash_{\texttt{PA}} \neg\varphi$, against the first incompleteness theorem.

REMARK 3. *By definition of the predicate $Theor(x)$, we have that*

$$Cons_{\texttt{PA}} \equiv \forall x \neg Dem(x, 0=1)$$

so that $Cons_{\texttt{PA}} \in \Pi_1$.

THEOREM 6 (Σ_1-completeness). PA *is Σ_1-complete w.r.t. the standard model $\mathcal{N}$, i.e. for any formula $\alpha \in \Sigma_1$, if $\mathcal{N} \vDash \alpha$, then $\vdash_{\texttt{PA}} \alpha$.*

Proof. At first, we show that PA is Δ_0-complete w.r.t. $\mathcal{N}$. Suppose α to be a Δ_0-formula such that $\mathcal{N} \vDash \alpha$, but $\nvdash_{\texttt{PA}} \alpha$. Since PA is Δ_0-syntactically complete, it would be $\vdash_{\texttt{PA}} \neg\alpha$ and so $\mathcal{N} \vDash \neg\alpha$.

We proceed by absurd again: let $\mathcal{N} \vDash \exists x\alpha(x)$, but $\nvdash_{\texttt{PA}} \exists \mathrm{x}\alpha(\mathtt{x})$. For $\mathcal{N} \vDash \exists x\alpha(x)$, there is an $n \in \mathbb{N}$ such that $\mathcal{N} \vDash \alpha(n)$. Since $\alpha(n) \in \Delta_0$, we can apply the just proved Δ_0-completeness and obtain $\vdash_{\texttt{PA}} \alpha(\mathtt{n})$. As a matter of logic, we finally obtain $\vdash_{\texttt{PA}} \exists \mathrm{x}\alpha(\mathtt{x})$ which contradicts our assumption that $\nvdash_{\texttt{PA}} \exists \mathrm{x}\alpha(\mathtt{x})$.

REMARK 4. *Σ_1-completeness cannot be generalised to the whole range of* PA *models: we can take as counterexamples all those non-standard models in which the unprovable proposition $\neg\varphi \in \Sigma_1$ turns out to be satisfied [7]. By looking at the previous demonstration, the key point consists in the application of the Δ_0-completeness which allows to pass from $\mathcal{N} \vDash \alpha(n)$ to $\vdash_{\texttt{PA}} \alpha(\mathtt{n})$. Consider a generic non-standard model $\mathcal{M}$: the formula $\exists x\alpha(x)$ may be instantiated by a non-standard element $a \in |\mathcal{M}| \backslash \mathbb{N}$. Yet in such a case we could not pass from $\mathcal{M} \vDash \alpha(a)$ to its syntactical side, since* PA *lacks the linguistic resources to express the formal counterpart of the element a. In other words, in case of non-standard models, the interpretation function is not surjective and this induces the failure of the Σ_1-completeness.*

COROLLARY 1. *If $\alpha \in \Pi_1$ is independent from* PA*, then $\mathcal{N} \vDash \alpha$. In particular, we have that $\mathcal{N} \vDash \varphi$ and $\mathcal{N} \vDash Cons_{\texttt{PA}}$.*

Proof. By the Σ_1-completeness, we obtain $\mathcal{N} \nvDash \neg\alpha$ from $\nvdash_{\texttt{PA}} \neg\alpha$, and so $\mathcal{N} \vDash \alpha$. Both the gödelian propositions φ and $Cons_{\texttt{PA}}$ instantiate the just explained case so that $\mathcal{N} \vDash \varphi$ and $\mathcal{N} \vDash Cons_{\texttt{PA}}$.

The following table summarises the technical results since now recalled:

	Δ_0	Σ_1	Π_1	higher levels
syntactical completeness	**yes**	no	no	no
semantical compl. w.r.t. $\mathcal{N}$	**yes**	**yes**	no	no
soundness w.r.t. $\mathcal{N}$	**yes**	**yes**	**yes**	**yes**

We conclude this section by recalling the Tarski's theorem about the undefinability of the truth within PA. While the incompleteness theorems underline the deductive limits of PA, the next result singles out an expressive limit of the theory.

DEFINITION 6. *(truth definition) An unary predicate* Tr(x) *is said to be truth definition for a theory* T *if, for any formula* α*, it is* $\vdash_{\mathtt{T}} \mathrm{Tr}(\ulcorner\alpha\urcorner) \leftrightarrow \alpha$.

THEOREM 7 (Tarski's theorem). *If* PA *is consistent, then it does not admit a truth definition.*

3. Gödel's Princeton Lectures: Footnote 21.

In this section, we provide a technical explanation of the footnote 21 appearing in the notes taken by Kleene and Rosser of the lectures that Gödel gave at the Institute for Advanced Study during the spring of 1934 [4]. Our primary concern is to isolate the original gödelian notion of arithmetical truth. As we will remark in the next section, the content of the footnote 21 strictly reflects what Gödel says at length later, in the paragraph 7 of the same notes, considering classical paradoxes.

In order to simplify our notation, we deal with 1-argument functions (any n-adic function can be in fact always reduced to an 1-adic one by a suitable encoding). According to the standard notation, we denote with PR the class of primitive recursive functions. We report below the claim of the representation theorem which plays a key role in Gödel's construction. It states that PA is sufficiently powerful to express the whole range of PR functions [3, 4, 6]. Formally:

THEOREM 8 (representation). *For each function* $f \in \mathtt{PR}$ *there is a formula* $\phi_f(x, y) \in \Sigma_1$*, such that, for any pair* $n, m \in \mathbb{N}$,

$$\mathcal{N} \models f(n) = m \Rightarrow \vdash_{\mathtt{PA}} \phi_f(\mathtt{n}, \mathtt{m}).$$

At the end of the proof, Gödel adds a footnote at Theorem 7, the number 21, where he remarks that:

> *[...] this proof is not necessary for the demonstration of the existence of undecidable arithmetic propositions in the system considered. For, if some recursive function were not "represented" by the corresponding formula [...], this would trivially imply the existence of undecidable propositions unless some wrong propositions on integers were demonstrable[1].*

In practice this remark amounts to stating the following:

THEOREM 9. *Suppose that the system* $\mathtt{PA}$ *is not able to "represent" the totality of* $\mathtt{PR}$ *functions; then, there exists a proposition* ϕ *independent from* $\mathtt{PA}$.

Proof. Suppose that the representation fails for a certain function $f \in \mathtt{PR}$ over a certain pair $a, b \in \mathbb{N}$, namely $\mathcal{N} \models f(a) = b$, but $\nvdash_{\mathtt{PA}} \phi_f(\mathtt{a}, \mathtt{b})$. By the Σ_1-completeness, we obtain $\mathcal{N} \not\models f(a) = b$ from $\nvdash_{\mathtt{PA}} \phi_f(\mathtt{a}, \mathtt{b})$ and so $\mathcal{N} \models f(a) \neq b$ against our assumption.

REMARK 5. *It is worth noting that the independence phenomenon induced by the failure of the representation is not related to the specific range of axioms which constitutes the deductive system at issue. As a matter of fact, nothing prevents the deductive hole*

$$\nvdash_{\mathtt{PA}} \phi_f(\mathtt{a}_i, \mathtt{b}_i)$$

from being limited to finitely many pairs of values a_i, b_i $(i = 1, \ldots, k)$*: in such a case, we could easily recover the syntactical completeness by adding k further axioms:*

$$\vdash_{\mathtt{PA}} \phi_f(\mathtt{a}_i, \mathtt{b}_i).$$

In principle, this kind of undecidability is not intrinsically insurmountable.

On the contrary, the undecidable proposition φ *constructed by Gödel involves the predicate* $Theor(x)$ *which is strictly depending on the set of axioms of* $\mathtt{PA}$*, so that it gives rise to an independence phenomenon which is intrinsically insurmountable. That is to say, we can consider the enriched formal system* $\mathtt{PA}' = \mathtt{PA} \cup \{\varphi\}$ *in which* φ *appears as an axiom and iterate Gödel's construction so as to produce another independent proposition* φ'*. Since* $\nvdash_{\mathtt{PA}'} \varphi'$ *and* $\vdash_{\mathtt{PA}'} \varphi$*, we clearly have* $\varphi' \neq \varphi$*, namely the deductive hole replicates itself into a new independent proposition.*

[1] K. Gödel. *On Undecidable Propositions of Formal Mathematical Systems.* In M. Davis *The Undecidable*, Raven Press, New York, 1965, p. 59.

4. Five Views of Truth

In this section, we briefly analyse the five fundamental arguments in favour of the truth of φ arose in the debate on incompleteness:

1. **The classical view**. The first incompleteness theorem (Theorem 4) states the independence of φ from $\mathtt{PA}$, i.e. $\nvdash_{\mathtt{PA}} \varphi$ and $\nvdash_{\mathtt{PA}} \neg\varphi$. In classical perspective, we know that either φ or $\neg\varphi$ must be true and so, in such an obvious way, Gödel shows the existence of a true yet unprovable proposition. Remark that in the present case we are not able to indicate which one between φ and $\neg\varphi$ is the true proposition. This argument expresses the platonistic point of view, according to which mathematical truth is prior to any kind of verification process, irrespective to any lack of information about this process.

2. **The autoreferential view**. On this view, φ is a true sentence, because φ says of itself that it is unprovable and it is indeed unprovable. Though informally, such an argument can be precised by looking at the proof of the first incompleteness theorem in which $\vdash_{\mathtt{PA}} \neg\varphi$ is deduced from the hypothesis by absurd that $\vdash_{\mathtt{PA}} \varphi$. This clearly contradicts the consistency of $\mathtt{PA}$ and so we are forced to conclude that φ cannot be provable [8].

3. **The extendend autoreferential view**. In [5], Longo has proposed to stress the role of second incompleteness theorem (Theorem 5) in order to precise the autoreferential argument. The second incompleteness theorem states that $\vdash_{\mathtt{PA}} \mathtt{Cons}_{\mathtt{PA}} \rightarrow \varphi$; from a semantical point of view $\mathcal{N} \vDash Cons_{\mathtt{PA}} \rightarrow \varphi$, and so, if $\mathcal{N} \vDash Cons_{\mathtt{PA}}$, then $\mathcal{N} \vDash \varphi$. In conclusion: assuming the consistency of $\mathtt{PA}$, we immediately have the truth of φ.

4. **The model-theoretical view**. As already seen, an easy consequence of the Σ_1-completness (Theorem 6) is that any formula $\alpha \in \Pi_1$ independent from $\mathtt{PA}$, turns out to be true in $\mathcal{N}$ (Corollary 1). Therefore, since $\varphi \in \Pi_1$, we have $\mathcal{N} \vDash \varphi$. This is just a refinement of the argument proposed in point 1: we still stress the excluded middle (see the proof of Theorem 6), but here the Σ_1-completeness enable us to pick up the true formula.

5. **The provability view.** The notion of truth coincides with the provability within $\mathtt{PA}$. Thus, a sentence is true just in case is derivable from Peano Axioms, and false just in case its negation is so

derivable. It follows that undecidable sentences are neither true nor false.

To pursue Point 1 is to make sense of incompleteness in terms of notions too much philosophically compromised to act at the same level with the other points. Briefly, we consider the remaining four points. Following the argument suggested by Longo, point 2 can be reduced to the number 3, or better, this latter gives a precise logical meaning to the autoreferential argument: $\mathcal{N} \models \varphi$ is obtained from $\mathcal{N} \models Cons_{\mathtt{PA}} \rightarrow \varphi$, by assuming $\mathcal{N} \models Cons_{\mathtt{PA}}$. We can perform one more step in the same direction, by remarking that it is not needed to assume $\mathcal{N} \models Cons_{\mathtt{PA}}$, since it is just a consequence of Corollary 1. In such a way, point 3 is in turn reduced to point 4. The argument proposed in point 4 is a real proof of $\mathcal{N} \models \varphi$ displayed within the specific resources of model theory. This leads us to trace point 4 back to point 5.

We conclude this section by remarking that Gödel's incompleteness results constitute a particular case in which the naive achievement of the truth (point 2) agrees with that one suggested by model-theoretical considerations (points 3 and 4). Nevertheless, there are no reasons to consider such an agreement to be generally valid and the naive argument may constitute a very misleading approach. Consider for instance the intuitionistic first order calculus LJ. The proposition $\psi \vee \neg\psi$ turns out to be independent from LJ and, in particular, the fact $\nvdash_{\mathtt{LJ}} \psi \vee \neg\psi$ is obtained by stressing a well-known argument by absurd. Now, the excluded middle should be considered as a true principle (I used it for proving the independence of $\psi \vee \neg\psi$!), but the model-theoretic version diverges from such an immediate intuition since Heyting algebras do not recognise $\psi \vee \neg\psi$ as a valid formula [2].

5. Conclusions

The truth of the gödelian proposition φ turns out to be necessarily *internal*. Because of its logical dependence to the Σ_1-completeness w.r.t. $\mathcal{N}$, the truth of φ *must* be confined to the range of the standard model of formal arithmetic (Remark 4) and, moreover, it comes at the end of a well-defined model-theoretic argument (Theorem 6 and Corollary 1).

Rereading Gödel's 1931 work in this light, we can clearly distinguish the awareness of the discrepancy between an inside and outside conception of truth and their respective connections with the concept of provability. Gödel, in fact, understands that an inside conception of truth would have once again involved the risk of the liar's paradox and it would have made a demonstration of the first incompleteness theorem in PA impossible. Although, in fact, semantic proofs of Gödel's theorems[2] are possible, their formalization in PA would require to define in PA the notion of truth for the statements of PA, which is excluded by Tarski's theorem on the indefinability of truth (Theorem 8). In contrast, a syntactic proof, as Gödel deftly showed, can be easily formalized in PA and this syntactic proof can only be linked to a local notion of truth which comes close to explaining the paradox without actually doing it, and which constitutes, as pointed out by Floyd & Putnam, a metaphysical statement which is distinct from the actual mathematical result (the undecidability).

Therefore, the substantial discrepancy between an inside and an outside conception of truth arises primarily from the need to distinguish between a notion that cannot be defined in PA without contradiction, and a notion, like provability in PA, which is semi-representable in PA. This point clearly emerges, for example, from reading the 48th note of Gödel's work of 1931:

> *As will be shown in Part II of this paper, the true reason for the incompleteness inherent in all formal systems of mathematics is that the formation of ever higher types can be continued into the transfinite (see Hilbert 1926, page 184), while in any formal system at most denumerably many of them are available. For it can be shown that the undecidable propositions constructed here become decidable whenever appropriate higher types are added (for example, the type w to the system P), An analogous situation prevails for the axiom system of set theory*[3].

The reference to appropriate higher types is precisely the means which enable to define the notion of truth for the PA and then decide φ which is unprovable in PA. The concept of truth itself is undefinable in PA because of Tarski's theorem, therefore φ must be decided in a context where we can express the semantic reasoning that *leads us to prove the truth of* φ.

Therefore, there are good reasons to disentangle truth from undecidability as well as to think that Gödel endorses the distinction between inside and outside views of truth.

[2] See R. Smullyan, *Gödel's Incompleteness Theorems*, Oxford University Press, 1992, ch. X.

[3] Gödel K., *Über formal unentscheidbare Sätze der Principia mathematica und verwandter Systeme*, Monatshefte für Mathematik und Physik, 1931, 38, pp. 173-198; also in Gödel K., *Collected Works,* I, Oxford: Oxford University Press, 1986, pp. 144-195, p. 18.

References

1. S. Feferman. *The impact of Gödel's incompleteness theorems on mathematics.* Notices American Mathematical Society 53 no. 4 (April 2006), 434-439.
2. PT J.-Y. Girard. *Proofs and Types.* Cambridge University Press, 1989.
3. K. Gödel. *On Formally Undecidable Propositions of* Principia Mathematica *and Related Systems.* In M. Davis *The Undecidable*, Raven Press, New York, 1965.
4. K. Gödel. *On Undecidable Propositions of Formal Mathematical Systems.* In M. Davis *The Undecidable*, Raven Press, New York, 1965.
5. G. Longo. *Reflections on Incompleteness* Lecture Notes in Computer Science, vol 2277 (Callaghan et al. eds), pp. 160 - 180, Springer, 2002.
6. E. Mendelson. *Inroduction to Mathematical Logic.* D. Van Nostrand Company – Princeton, New Jersey – 1964.
7. R. Kaye. *Models of Peano Arithmetic.* Clarendon Press, Oxford, 1991.
8. J. Peregrin. *Gödel, Truth & Proof.* Journal of Physics: Conference Series 82, 2007.
9. C. Smoryńsky. *The Incompleteness Theorems.* Handbook of Mathematical Logic, North Holland, 1977.
10. J. Floyd, H. Putnam. *A note on Wittgenstein's 'Notorious paragraph' about the Gödel theorem.*Journal of philosophy, 97, 2000.
11. R. Smullyan. *Gödel's Incompleteness Theorems.* Oxford University Press, 1992.