Is G true by Gödel's theorem?

Virgil Drăghici
Philosophy Dept.
Babeş-Bolyai University
Cluj-Napoca, Romania
E-mail: vidra007@gmail.com

Abstract:
Two philosophical arguments, e.g. that the meaning of an expression transcends its use and that the human arithmetical thinking is not entirely algorithmic (the argument Lucas/Penrose) base their theses on Gödel's first incompleteness theorem. But both in these arguments and in some of their criticisms the word "true" is often used ambiguous: it swings between a licit metamathematical use and an illicit transfer of it in a formal system. The aim of this paper is to show the way these arguments are connected, via G-type sentences (sect 2), and how do we argue that the sentence G, albeit unprovable in PA, is true, by using non-conservative extensions of PA with reflections (sect 3). And this without any illicit use of "true".

Keywords: first incompleteness theorem, G-type sentences, Peano Arithmetic, provability, truth, reflection principles.

1. Preliminary

Gödel's first incompleteness theorem[1] is a fundamental result of mathematical logic. But its philosophical relevance is still a matter of controversy. Two famous theses are based on this theorem: the idea that

[1] Cf. K. Gödel [1].

meaning of an expression transcends its use, and the idea that our arithmetical understanding is not entirely algorithmic (the antimechanist view or the argument Lucas/Penrose). In both cases, roughly speaking, the conclusion is the same: by Gödel's theorem there are true arithmetical sentences beyond any systematic use of them or algorithmic means to show them to be true. This is the case with G-type sentences. But is this really so?

Moreover, in some criticisms of these views some exceeding use of the word "true" related to such sentences occurs. Let us take two notable examples.

In a comment on the antimechanist view Putnam[2] says that

> [g]iven an arbitrary machine T, all I can do is find a proposition [G] such that I can prove:
>
> (3) If T is consistent, [G] is true,
>
> where [G] is undecidable by T if T is in fact consistent. However, *T can perfectly well prove (3) too!* (*emphasis added*).

In referring to the question whether Gödel's result can be an argument against the idea that the meaning of an arithmetical concept outruns its use in a formal system, M. Dummett[3] wrote:

> Considered as an argument to a hypothetical conclusion – that *if* the system is consistent, then $\forall x A(x)$ is true – this reasoning can of course be formalized in the system.

Both quotations contain a misuse of "true", for both the sentence (3) in Putnam quotation and the conditional mentioned by Dummett, *accurately formalized*, are implications of the form $Con(T) \supset Tr(\ulcorner G \urcorner)$ containing the truth predicate $Tr(x)$, and by Tarskis's theorem an insertion in T of a truth predicate will make it inconsistent.

Let us see firstly how the above arguments are connected.

---

[2] Cf. H. Putnam [1], 366.

[3] Cf. M. Dummett [1], 192.

2. G-type sentences

2.1. Gödel's original result

In more usual terms, this theorem can be stated as follows:

Th G1. a) *If* S *is consistent, then* G *is not provable in* S.

b) *If* S *is* ω-*consistent, then* ¬G *is not provable in* S,

where S is an adequate first-order formalization of Peano Arithmetic (PA).

If *Pf*(*x,y*) is the primitive recursive relation "*x* is a proof of *y*" and π(x,y) is the formula expressing it in PA, then by one application of DL to the formula ∀x¬π(x,y) we obtain

$$\text{PA}\vdash \text{G}\equiv\forall x\neg\pi(x,\ulcorner \text{G}\urcorner), \qquad (1)$$

where ⌜G⌝ is the formal name of G. That is, G is a sentence provably equivalent to a sentence asserting that G is not provable. Equivalent formulations of (1) are PA⊢G≡¬∃xπ(x,⌜G⌝), PA⊢G≡¬∃xπ(x,***g***), where ***g*** is the numeral for the Gödel number of G, and PA⊢G≡¬Bew(⌜G⌝), where Bew(y) is the provability predicate ∃xπ(x,y). As can be seen, by Th G1a) what is really proved is a *conditional* assertion, without any reference to the truth of G.

2.2. Kleene's generalization of Th G1a)

By his generalization[4], if S is a formal system in which the predicate $(y)\overline{T}_1(x,x,y)$ is expressed by α(x), then a number *k* can be found such that

> If S is correct for $(y)\overline{T}_1(x,x,y)$, then $(y)\overline{T}_1(k,k,y)$ and [$\nvdash$ α(***k***)]; i.e. the proposition $(y)\overline{T}_1(k,k,y)$ is true, but the formula [α(***k***)] expressing it is unprovable.

In this quotation the word "true" occurs after the semicolon, in an explanatory sentence with respect to the preceding one. So "$(y)\overline{T}_1(k,k,y)$" and ""$(y)\overline{T}_1(k,k,y)$" is true" means the same thing. Remember that in Kleene's formulation $T_1(z,x,y)$ is a primitive recursive predicate. If we take

[4] Cf. S.C. Kleene [1], Theorem XIII (Part II), 303.

the formula θ(z,x,y) as expressing it in PA, then the formula expressing the relation $(y)\overline{T}_1(k,k,y)$ (Kleene's intuitive predicate) is just the G-type sentence ∀y¬θ(***k***,***k***,y), a formula not provable in S.

As can be seen, what is proved in Kleene's formulation of Gödel's result is also a *conditional*, whose hypothesis is the correctness of S. Only under such an assumption we can derive either G or "G is true" but as we show bellow, *not within* PA. So, once more, neither G nor "G" *is true* follows by Gödel's theorem.

2.3. Turing's version of Th G1a)

In computational terms, Th G1a) has the following form: *If A*(*q*,*n*) *is a sound procedure for ascertaining the non-halting of the computation* $C_q(n)$, *then it is incomplete.* More exactly, by using Cantor's diagonal method it can be argued that for some *k* the computation $C_k(k)$ does *not* halt and *A*(*k*,*k*) cannot halt either. So, *if A*(*q*,*n*) *is sound*, then it is incomplete.

As is well-known, every formal system can be recast as a theorem-proving machine and vice-versa. Indeed, Turing's version is similar to Kleene's generalization of Gödel's theorem. For Kleene's T-predicate $T_1(z,x,y)$ does admit of a computational "translation": *y* is the Gödel number of a computation at input *x* on the Turing machine C_z. So, in terms of the above theorem, $(y)\overline{T}_1(k,k,y)$ says simply that $C_k(k)$ does not stop. And in terms of formal systems, "$C_k(k)$ does not stop" is just the G-type formula mentioned above: ∀y¬θ(***k***,***k***,y).

A simple inspection of these forms of Gödel's result shows that in all cases what is proved is a *conditional* assertion of the form "If..., then ___". So, by these theorems we have no argument for the truth of G. But, of course, between the formulation of Th G1a) and Kleene's formulation there is a difference. It follows from what can and what cannot be proved within PA.

Very concisely, the following holds:

$$\text{PA}\vdash \text{Con(PA)}\equiv\neg\text{Bew}(\ulcorner\bot\urcorner)\equiv G\equiv\neg\text{Bew}(\ulcorner G\urcorner)$$

So, Con(PA)⊃G is a theorem of PA, but not Con(PA)⊃Tr(⌜G⌝), as Putnam and Dummett refer (incorrectly) as being provable in PA or by a Turing machine. However, G is true. But how can that be argued?

3. How do we know that G is true?

As the preceding considerations show in order to derive either G or "G" *is true* we need an argument for the antecedent of the implication Con(PA)⊃G, properly extending PA. We can do that either *metamathematically* (informally) or *formally*, by inserting the requested condition in a formal system.

A concise form of the first way was given by Dummett[5] in the following terms. As we saw, G is (equivalent to) the sentence $\forall x \neg\pi(x,\boldsymbol{g})$, where $\pi(x,\boldsymbol{g})$ is a primitive recursive (also decidable) formula of PA. Hence every instance of this formula is provable in PA. The truth of G follows therefore by the following steps: a) being provable, all sentences $\neg\pi(\boldsymbol{0},\boldsymbol{g})$, $\neg\pi(\boldsymbol{1},\boldsymbol{g})$, $\neg\pi(\boldsymbol{2},\boldsymbol{g})$,..., are *true* (in the standard model M), and b) by the semantics of "∀" it follows that $\forall x \neg\pi(x,\boldsymbol{g})$ is also *true* in M.

The key step is a) and it expresses the idea of *soundness of* PA: *every provable formula of* PA *is true in* M.

Both Dummett's metamathematical argument and Kleene's metamathematical formulation of Gödel's theorem suggest the way G's truth can be *formally* derived: a system of arithmetic in which Con(PA) or *an equivalent sentence* is actually a formula of this system. There are many ways to construct such a system, depending on what we intend to formalize and to prove. Let us see concisely some of them.

If what we want is just a derivation of G, then a non-conservative[6] extension of PA of the following form is needed: PA^*=PA+Con(PA). This is not attractive for the simple reason that a derivation of G in PA^* presupposes the complicated formal proof of Con(PA)⊃G. Alternatively, we can extend PA with a reflection principle (asserting correctness of PA, as Kleene's formulation requires), but *without* any use of the truth predicate

[5] Cf. M. Dummett [1], 191.

[6] T^* is conservative over *T* if any formula of *T*, provable in T^*, is also provable in *T*.

Tr(x), i.e. PA^*=PA+Refl. Indeed, such a reflection is available, it is an uniform reflection principle[7]:

URefl: $\forall x Bew(\ulcorner\alpha(\boldsymbol{x})\urcorner)\supset\forall x\alpha(x)$; α has only x free.

A result of C. Smorynski[8] establishes that if $\alpha(x)$ in URefl is a π_1-formula (like the sentence G), then the following equivalence holds: Con(PA)≡URefl. And, finally, if we consider the fact that G has the form $\forall x\neg\pi(x,\boldsymbol{g})$, in which $\neg\pi(x,\boldsymbol{g})$ is primitive recursive, then URefl can be weakened to primitive recursive formulas, the Smorynski's equivalence being preserved. Therefore PA^*=PA+URefl, where URefl is so restricted, allows a derivation of G[9].

Let us end with a remark. If what we want is a derivation of $Tr(\ulcorner G\urcorner)$, in an non-trivial fashion, then we need an extension of PA with a Tarskian theory of satisfaction (truth), i.e. PA^*=PA+*Sat*, where *Sat* is a theory formulated in $L^*_{PA} = L_{PA}$+Sat(x,y), containing the axioms for satisfaction predicate Sat(x,y). As Tarski showed[10], such an extension proves that *any provable formula of* PA *is true*, i.e. just the soundness (reflection) of PA. Of course, what is obtained is not a complete theory, for it has its own Gödel sentence G^* not provable in PA^*. Moreover, G is provable in PA^* and assertible in PA^* only if PA^* is a reliable system of proof. Hence the soundness of PA^* is still an assumption for the assertibility of G.

4. Conclusion

The truth of the sentence G does not follow in any way from Gödel's theorem. Its truth or its assertibility can be shown either *metamathematically*, by an informal argument transcending PA, or *formally*, by a proof in a non-conservative extension of PA with reflection

---

7 Cf. S. Feferman [1], section 2(d).

8 Cf. C. Smorynski [1], Theorem 4.1.4.

9 In a very interesting paper (cf. Tennant [1]) Tennant used such a restriction in order to argue that it is the minimum condition on an extension of PA in which G is provable. Moreover, such an extension makes a conversion of "semantical argument" for the truth of G (requiring a 'thick' notion of truth) into a syntactical one, perfectly compatible with a non-constructive version of deflationism (requiring a 'thin' notion of truth).

10 Cf. A. Tarski [1], Theorem 5; comp. and S. Feferman [2], 16, Theorem 2.5.3.

principles[11]. And this fact determines a reassessment of both arguments based on this theorem and of some of their criticisms.

References

Dummett, M., [1] "The philosophical significance of Gödel's theorem (1963)", in *Truth and Other Enigmas*, London, Duckworth, 1978, 186-201.

Feferman, S., [1] "Transfinite recursive progressions of axiomatic theories", *Journal of Symbolic Logic*, 27, 1962, 259-316.

[2] "Reflecting on incomplenetess", *Journal of Symbolic Logic,* 56, 1991, 1-49.

Gödel, K., [1] "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I", *Monatsh. Math. Phys*., 38, 1931, 173-198.

Kleene, S. C., [1] *Introduction to Metamathematics*, fourth edition, North-Holland Publishing Co., Amsterdam, 1964.

Putnam, H., [1]"Minds and machines", in *Philosophical Papers* 2, Cambridge University Press, 1975, 362-385.

Smorynski, C., [1] "The incompleteness theorems", in *Handbook of Mathematical Logic*, J. Barwise ed., North-Holland Publishing Company, Amsterdam, 1977, 821-865.

Tarski, A., [1] "The concept of truth in formalized languages", in *Logic, Semantics, Metamathematics: Papers by Alfred Tarski from* 1922-1938, Oxford: Clarendon Press, 1956, 152-278. Originally published as "Der Wahrheitsbegriff in den formalisierten Sprachen", *Studia Philosophia* I, 1936.

Tennant, N., [1] “Deflationism and the Gödel Phenomena”, *Mind*, 111, 2002, 551-582.

[11] Or, equivalently, by using a weaker form of the ω-rule whose only premise is the expression $PA \vdash \forall x Bew(\ulcorner \neg\pi(\boldsymbol{x},\boldsymbol{g})\urcorner)$.