teorema
Vol. XXXV/1, 2016, pp. 5-12
ISSN: 0210-1602
[BIBLID 0210-1602 (2016) 35:1; pp. 5-12]

On Reliability of the Meta-Mathematical Notions Defined by Gödel's Coding Method

Ehsan Siavashi

RESUMEN

Se ha dado por sentado que el método de codificación de Gödel es un procedimiento fiable para definir nociones metamatemáticas en cualquier extensión de la aritmética de Robinson (Q). Sin embargo, puede mostrarse que algunas fórmulas definidas de acuerdo con este método e interpretadas como el predicado de demostrabilidad, el predicado de indemostrabilidad o la oración de consistencia, no logran satisfacer algunos requisitos. Por ejemplo, se sabe que algunas extensiones de la aritmética de Robinson demuestran sus propias oraciones canónicas de *in*consistencia, mientras que estas son efectivamente consistentes. Una respuesta común a este problema es que esas teorías son incorrectas y las teorías erróneas podrían demostrar cosas erróneas como, por ejemplo, su propia inconsistencia. Este artículo argumentará por qué tales respuestas no son totalmente convincentes. Al final, el artículo sugiere una lectura del primer y segundo teoremas de incompletud que está libre de esas interpretaciones.

PALABRAS CLAVE: *teoremas de Gödel, interpretación, metamatemática, definibilidad.*

ABSTRACT

It has been taken for granted that Gödel's coding method is a reliable method for defining meta-mathematical notions in every extension of Robinson Arithmetic (Q). However, it could be shown that some formulas defined by the method and interpreted as provability predicate, unprovability predicate or consistency sentence, fail to satisfy some requirements. For example, it is known that some extensions of Robinson Arithmetic prove their own canonical *in*consistency sentences, while they are actually consistent. A common response to this problem is that those theories are unsound, and wrong theories might prove wrong things such as their own inconsistency. However, the paper will argue why such responses are not totally convincing. At the end, the paper suggests a reading of the first and the second incompleteness theorems which is free from such interpretations.

KEYWORDS: *Godel's theorems*, *Interpretation*, *Meta-Mathematics*, *Definability*.

I. INTRODUCTION

Gödelian coding method is a method which enables extensions of Robinson Arithmetic[1] (Q) to "talk" about their own meta-theoretical

properties. However, this doesn't mean that whatever they "say" about themselves is also reliable. Although the way that Gödel codes the meta-mathematical notions sounds pretty safe and convincing, it is disputable whether or not there is enough connection between those formulas and their related meta-mathematical concepts. For instance, assume that T is an extension of Q and $\sim CON_T$ is the sentence defined by Gödel's method and interpreted as the *in*consistency sentence of T. It is known that:

(I) $T \vdash \sim CON_T$ doesn't imply that T is *in*consistent.

To see why (I) holds, consider the theory $PA^* = PA + \{\sim CON_{PA}\}$. Consistency of PA implies consistency of PA^*.[2] In fact, if $PA^* \vdash \bot$, by deduction theorem, $PA \vdash \sim CON_{PA} \rightarrow \bot$, which contradicts the second incompleteness theorem. Therefore, PA^* is consistent. That being said, interestingly enough, PA^* proves its own *in*consistency sentence (for a proof, please see [Smith (2007), pp. 225])[3]. In summary:

(II) $PA^* \vdash \sim CON_{PA^*}$, but it is consistent.

A similar problem arises if we consider the canonical provability predicate of PA^*. One might propose the following condition as a necessary condition for every provability predicate:

(III) If $T \vdash Prov_T(\text{'}s\text{'})$, then $T \vdash s$.

Yet, $Prov_{PA^*}$ does not satisfy (III). As was mentioned above, PA^* proves its own inconsistency sentence $\sim CON_{PA^*}$ which is actually the sentence: $Prov_{PA^*}(\text{'}\bot\text{'})$. And as was mentioned before, PA^* is consistent. Therefore:

(IV) $PA^* \vdash Prov_{PA^*}(\text{'}\bot\text{'})$, but $PA^* \nvdash \bot$.

which is a counterexample for (III).

On the face of these problems, logicians usually appeal to the distinction between provability and truth. For example, that CON_{PA^*} is provable in PA^* does not mean that it is also true. In fact, since PA^* is a false theory (because its axiom $\sim CON_{PA}$ is false), it is not surprising that it proves false things such as $\sim CON_{PA^*}$. Peter Smith explains this view in this way:

> What are we to make of this apparent absurdity? Well, giving the language of PA^* its standard arithmetical interpretation, the theory is just wrong in what it says about its inconsistency! But on reflection that shouldn't be

> much of a surprise. Believing, as we no doubt do, that *PA* is consistent, we already know that the theory *PA** gets things wrong right at the outset, since its axioms aren't all true. So *PA** doesn't actually prove (establish-as-true) its own inconsistency, since we don't accept the theory as correct on the standard interpretation [Smith (2007), pp. 225].

In the case of the predicate $Prov_{PA^*}$, a similar response is usually given. The claim is that (III) is not actually a provability condition. That a system proves its own inconsistency, doesn't mean that it is actually inconsistent. In fact, this is why (III) is not one of the Hilbert-Bernays-Löb conditions for provability. According to this view, $Prov_{PA^*}$ *is* actually a provability predicate for *PA**. The fact that it does not satisfy (III) is caused by *PA** being an unsound theory.

II. IS THE RESPONSE CONVINCING?

The response, however, does not seem to be convincing. One can attack such claims in several ways. But, the main argument against it is the fact that such problems are not limited to the unsound extensions of *Q*. For instance, consider *PA* and imagine a possible world where, after arithmetizing the meta-theory of *PA* and before discovering the second incompleteness theorem, Gödel erroneously provides a (wrong) proof for $PA \vdash CON_{PA}$. Would then he claim that he has proved consistency of *PA*? No. A proof for CON_{PA} in *PA* would not mean that *PA* is consistent, because every *in*consistent theory (which can express its own syntax) can also prove its own consistency sentence. As Raymond Smullyan mentions, not paying attention to this point has lead to some misunderstandings:

> We have seen such irresponsible statements as, 'By Gödel's second theorem, we can never know whether or not arithmetic is consistent.' Rubbish! To see how silly this is, suppose it had turned out that the sentence CON_{PA} were provable in *PA*—or,... suppose we consider a system that can prove its own consistency. Would that be any grounds for trusting the consistency of the system? Of course not! If the system were inconsistent, then it could prove every sentence—including the statement of its own consistency! To trust the consistency of a system on the grounds that it can prove its own consistency is as foolish as trusting a person's veracity on the grounds that he claims that he never lies [Smullyan (1992), pp.109].

The assumptions that Smullyan makes here are not unrealistic at all. Indeed, Solomon Feferman has showed that we can define *extensionally*

consistency statements of *PA* such that they *are* actually provable in *PA*. However, as he says: "rather than contradicting Gödel's second underivability theorem, ... [the theorems] show the importance of a precise method of dealing with consistency statements".[Feferman (1960), p. 69] Anyways, the fact that these consistency sentences are provable in *PA* does not mean that *PA* proves its own consistency, neither it adds any strength to our believe that *PA* is consistent. In fact, even without knowing about the second incompleteness theorem, one can see that:

(V) $T \vdash CON_T$ does *not* imply that *T* is consistent.

Unlike (I), it cannot be claimed that this "absurdity" is due to the unsoundness of *T*, because we haven't made any assumption about soundness or unsoundness of the system here. In summary, (I) and (V) show that, whether a theory *T* proves CON_T or $\sim CON_T$ (by itself) says nothing about consistency or inconsistency of *T*. Merely the fact that CON_T is defined by Gödelian coding method is irrelevant and cannot ensure us that there is any relation between CON_T ($\sim CON_T$) and consistency (inconsistency) of *T*.

The second incompleteness theorem makes the situation even worth. By the second theorem, we know that if *T* is a presumably *sound* extension of *Q* such as *PA*, then:

(VI) $T \vdash CON_T$, implies that *T* is in fact *in*consistent.

That is, if a sound theory like *PA* proves its own consistency, then such a proof not only cannot guarantee consistency of the theory, but also (even worth) shows that the theory is in fact *in*consistent. But, how then CON_{PA} can be a consistency sentence? Since CON_{PA} is equivalent to $\sim Prov_{PA}(\text{`}0=1\text{'})$, we can instead ask: how we know that $\sim Prov_{PA}$ is an unprovability predicate? Well, if $\sim Prov_{PA}$ is truly an unprovability predicate of *PA*, then $\sim Prov_{PA}(\text{`}0=1\text{'})$ must be true, otherwise our defined predicate, $\sim Prov_{PA}$, is not inclusive! Assuming that *PA* is sound, if the theory could prove $\sim Prov_{PA}(\text{`}0=1\text{'})$, we would have a reason to think that *0=1* is in the extension set of $\sim Prov_{PA}$. But, since we cannot prove it, the question remains unsettled. The usual response to this question is that *0=1 is* really in the extension set of $\sim Prov_{PA}$, but we simply just cannot prove it. My objection to this view is that it presupposes what it is supposed to show, that is, it presupposes that $\sim Prov_{PA}$ is really the unprovability pred-

icate of *PA*. But, $\sim Prov_{PA}$ is a predicate defined inside the theory and has no meaning outside it. If the theory doesn't know whether $\sim Prov_{PA}$*('0=1')* is true or false, we don't know either. In other words, the incompleteness of *PA* suggests that the defined predicate $\sim Prov_{PA}$ is not *inclusive*. To better see the point, let us use an analogy. Suppose that you have a theory about integers which contains a predicate named *Prime*. At some point, you realize that if *Prime*(*'879,190,747'*) is derivable from your theory, then 879,190,747 is not actually a prime number. But, on the other hand, you know that 879,190,747 *is* actually a prime number. You have two options to settle this absurdity: either reject the soundness of your theory, or reject the idea that *Prime* is the right predicate for prime numbers. Saying that *879,190,747* is in the extension set of *Prime* but we just cannot prove it is absurd, because the *Prime* predicate has its meaning inside the theory. Appealing to the incompleteness of your theory does not help, because, this just means that the defined predicate *Prime* is not inclusive!

III. What is the Cause of these Antinomies?

Since Gödel's incompleteness theorems were published in 1931, we have been encountered with many negative formal results regarding systems of arithmetic.[4] Furthermore, if we try to carry out philosophical interpretations of these formal results, we will be even more unsatisfied. George Boolos after discussing four controversial results of Löb's theorem asks:

> …it seems wholly bizarre that the statement that if *S* is provable, then *S* is true is not itself provable, in general. For isn't it perfectly obvious, for any *S*, that *S* is true if provable? Why we are bothering with *PA* if its theorems are false? And how could any such (apparently) obvious truth not be provable? [Boolos (1993), pp.55].

These antinomies have recently raised some discussions among mathematicians as well. Some have gone so far as to suggest that *PA* is inconsistent. In 2010, Fields medal laureate, Vladimir Voevodsky, claimed that consistency of *PA* is an open problem and more likely its answer is negative. In a lecture at Institute of Advanced Studies, he claimed that:

> …the correct interpretation of Gödel's second incompleteness theorem is that it provides a step toward the proof of inconsistency of many formal theories and in particular of the 'first order arithmetic [Voevodsky (2010)].

One year later, in September 2011, Edward Nelson claimed that he has proved inconsistency of *PA* [Nelson (2011)], though he later on retracted his claim due to a mistake in his proof found by Terry Tao.

None of the antinomies about *PA* provides enough evidence for rejecting consistency of *PA*. Instead of rejecting consistency of *PA*, I suggest that we must reject reliability of Gödelian coding method in defining meta-mathematical notions and reliability of the standard interpretations of those formulas. In my view, most of the above mentioned problems root in relying too much on our interpretations. For example, the second incompleteness theorem is sometimes called "unprovability of consistency" theorem (e.g. in [Boolos et al. (2007), pp. 232-243] and [Smith (2007), pp. 239-245]). This naming is certainly misleading, because it suggests that if it was possible to prove $PA \vdash CON_{PA}$, then consistency of *PA* was provable. But, as was pointed out by Smullyan, this is not the case. The sentence CON_{PA} either is not a consistency sentence or it is but in a weak unreliable sense.

IV. GÖDEL'S RESULTS

The view that I have proposed here, might cause some worries: if Gödel's method is not actually a reliable method for defining the meta-theoretical notions inside extensions of *Q*, how should we understand the first and the second incompleteness theorems? This is a worry, because sentences such as CON_{PA} and *G* (Gödel's sentence) which appear in these theorems are both defined by Gödel's method, and it seems like interpreting these sentences in this specific way plays an important role in achieving Gödel's results. In response to this worry, I will argue that we can hold all of the fascinating and important results of Gödel's theorems without such interpretations. What we will lose is confusion.

The First Incompleteness Theorem: The most important result of the first incompleteness theorem is that, assuming *PA* is consistent, it is incomplete. This is proved by introducing the Gödel sentence *G*. Gödel shows that neither *G* nor $\neg G$ is provable in *PA*.[5] At this point, people usually start giving a "meta-theoretical argument". Interpreting *G* as a sentence which says "I am not provable", it follows that *G* is indeed true. Since *PA* cannot prove this true sentence, it is incomplete.

However, we do not need to appeal to such meta-theoretical arguments to see that *PA* is incomplete. In order to show that *PA* is incom-

plete (if consistent), we just need to know that neither G nor $\neg G$ is provable. Since either G or $\neg G$ is true, there is a truth about arithmetic that is not provable. Therefore, by the first incompleteness theorem and without appealing to any kind of meta-theoretical argument or interpretation, we can get the result that: if *PA* is consistent, it is incomplete.

The Second Incompleteness Theorem: In my view, while the first incompleteness theorem introduces the independent sentence G, the second incompleteness theorem introduces another independent sentence relative to *PA* that is CON_{PA}. We know that these two sentences are equivalent and in fact: $PA\vdash G\leftrightarrow CON_{PA}$. A better and safer reading of the second theorem is to say: "if *PA* proves a specific sentence of itself represented by CON_{PA}, then *PA* is inconsistent".

Department of Computer Science
Texas Tech University
9th and Canton Avenue
Lubbock, TX 79409 USA
E-mail: ehsan.siavashi@ttu.edu

NOTES

[1] Although Gödel applied his coding method on Peano's Arithmetic, the method is actually applicable on some significantly weaker systems as well. Robinson Arithmetic or *Q* is known to be the weakest arithmetic theory that can "talk" about its syntax via Gödel's coding method. For a discussion on *Q* and its axioms please see chapter 10 of [Smith (2007), pp.62-71].

[2] However, *PA** is ω-inconsistent.

[3] It follows from the lemma $PA\vdash\sim CON_{T1}\rightarrow\sim CON_{T2}$ where *T1* and *T2* are two p.r. axiomatized theories and *T1* is a sub-theory of *T2*.

[4] Here are some examples: if Peano Arithmetic is consistent then it is incomplete (Gödel, 1931); this consistency is unprovable inside *PA* [Gödel (1931)]; consistency of no sentence s is provable in *PA*, even when s is a theorem of *PA* (i.e. no sentence in the form of $\sim\Box\sim s$ is a theorem of the provability logic); truth predicate is not definable *in PA* [Tarski (1933)]; if s is not provable in *PA*, then $Prov_{PA}(\text{`}s\text{'})\rightarrow s$ is not provable in *PA* [Löb (1955)]; the presumably wrong theory $PA+\{\neg CON_{PA}\}$ is interpretable in *PA*, while the right theory $PA+\{CON_{PA}\}$ is not interpretable in *PA* [Feferman (1960)] etc.

[5] To be historically more accurate, it should be mentioned that Gödel himself made the stronger assumption that *PA* is ω-consistent. It was Barkley Rosser who showed that the plain consistency assumption is enough [Rosser (1936)].

REFERENCES

BOOLOS, G. S. (1993), *The Logic of Provability*, Cambridge, Cambridge University Press.

BOOLOS, G. S., BURGES, J. and JEFFREY, R. (2007), *Computability and Logic*, Cambridge, Cambridge University Press.

FEFERMAN, S. (1960), 'Arithmetization of Metamathematics in a General Setting', *Fundamenta Mathematicae*, vol 49, pp. 35-92.

GÖDEL, K. (1931), 'On Formally Undecidable Propositions of Principia Mathematica and Related Systems'; (German:) 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38, pp. 173-198. Reprinted and translated in English by Dover Publications, 1992.

LOB, M. H. (1955), "Solution of a Problem of Leon Henkin"; *The Journal of Symbolic Logic,* vol. 20(2), pp. 115-118.

NELSON, E. (2011), link: *http://www.cs.nyu.edu/pipermail/fom/2011-September/ 015816.html*

ROSSER, B. (1936), "Extensions of some theorems of Gödel and Church", *Journal of Symbolic Logic*, vol. 1, pp. 87-91.

SMITH, P. (2007), *An Introduction to Gödel's Theorems*, Cambridge, Cambridge University Press.

SMULLYAN, R. M. (1992), *Gödel's Incompleteness Theorems*, Oxford, Oxford University Press.

TARSKI, A. (1953), *Undecidable Theories*, Dordrecht, North Holland Publishing Company.

VOEVODSKY, V. (2010), "What if the Current Foundations of Mathematics are Inconsistent?" Lecture at the celebration of the 80s anniversary of the IAS, Princeton, New Jersey, September, 25, 2010. link: *https://video.ias.edu/voevodsky-80th*.