Notre Dame Journal of Formal Logic
Volume IX, Number 1, January 1968

A PHILOSOPHICAL REMARK ON GÖDEL'S UNPROVABILITY OF CONSITENCY PROOF

FRANCESCA RIVETTI BARBÒ

This paper is concerned with the proof of theorem XI of Gödel's monograph of 1931.[1] The aim is to show that in this proof a G-number is:

1) introduced *as representing* the statement expressed by the associated formula;

2) then considered *as not representing it*;

this change being indispensable to the validity of Gödel's demonstration.

Therefore this proof does not meet the fundamental logico-philosophical canon: in any argument each term has always to be taken in the same way—and thus each word has to be used to mean or to represent (through arithmetization of syntax) the same entity—. Hence, from this philosophical point of view, the proof does not seem to be valid. To the contrary if the diagonal method (or anyhow the arithmetization of syntax) legitimates the above change, then, from such a point of view, the objection, I will expound, could not be raised (see below, § 7).

§1. I recall the main features of the proof. On the basis of the proof of theorem VI: if **c** is consistent, i.e., if Wid(**c**)[2], then 17 Gen **r** is not **c**-*provable* i.e.

$$\alpha)\quad \overline{\text{Bew}_{\mathbf{c}}}(17\ \text{Gen}\ \mathbf{r})\ ;$$

that is

$$(1)\quad \text{Wid}(\mathbf{c}) \rightarrow \overline{\text{Bew}_{\mathbf{c}}}(17\ \text{Gen}\ \mathbf{r})\quad \text{(Gödel's 23)}\ .$$

From (1) we derive[3]

$$(2)\quad \text{Wid}(\mathbf{c}) \rightarrow (\mathbf{x})\,\dot{\mathbf{Q}}(\mathbf{x},\mathbf{p})\qquad \text{(Gödel's 24)}$$

Concepts and assertion so far employed are expressible (or provable) in *P*; therefore, **w** being the G-number of the formula expressing Wid(**c**) and 17 Gen **r** the one of (**x**)**Q**(**x**,**p**), the G-number of the formula expressing (2) is

Received February 4, 1966

$$\mathbf{w}\ \mathrm{Imp}(17\ \mathrm{Gen}\ \mathbf{r})\ ,$$

which is *provable* (in *P*) and, a fortiori, **c**-*provable*. Thus, if **w** were **c**-*provable*, i.e. if $\mathrm{Bew}_{\mathbf{c}}(\mathbf{w})$, then also 17 Gen **r** would be **c**-*provable*, i.e.:

β) $\mathrm{Bew}_{\mathbf{c}}(17\ \mathrm{Gen}\ \mathbf{r})$

Hence, on the basis of (1): if **w** is **c**-*provable*, then **c** is not consistent. As **w** is the G-number of the formula w[4] expressing in c the arithmetical proposition Wid(**c**) equivalent to the metamathematical one "c is consistent"[5], no equivalent of this last statement can be proved in c, it being assumed that c is consistent. Thus the proof is based on the contradiction between the propositions expressed by α and by β[6], following respectively from Wid(**c**) and from **w** being **c**-*provable*.

§2. I will demonstrate that, if the proposition expressed by β follows from a proposition, $\mathrm{Bew}_{\mathbf{c}}(\mathbf{w})$, equivalent to "Wid(**c**) is provable in c", then: the occurrence of '17 Gen **r**'[5] in β stands for a number (coordinated to a formula) representing an entity different from the entity to which '17 Gen **r**' refers when occurring in α, granted that words and phrases are always used in the same way. Thus the proposition expressed by β does not contradict the one expressed by α. (See §4b and §4d; for the associated formulas and corresponding metamathematical concepts see §5c).

Hence the proposition expressed by β contradicts the one expressed by α if and only if, at the contrary, we obliterate that the occurrence of '17 Gen **r**' in β stands for such an entity, as only in this case the entity spoken about in both is the same. (See §4c; for the corresponding formulas and metamathematical concepts see §5d). Otherwise the contradiction would collapse, and therefore the proof too.

§3. On the basis of **w** Imp (17 Gen **r**) being **c**-*provable* we can assert that if $\mathrm{Bew}_{\mathbf{c}}(\mathbf{w})$ then also $\mathrm{Bew}_{\mathbf{c}}(17\ \mathrm{Gen}\ \mathbf{r})$ in so far as **w** Imp (17 Gen **r**) represents an implication; but this implication holds if and only if:

a) the formula represented by **w** (say w) expresses Wid (**c**);
b) the formula represented by 17 Gen **r** (say g) expresses $\overline{\mathrm{Bew}_{\mathbf{c}}}(17\ \mathrm{Gen}\ \mathbf{r})$, and $(\mathbf{x})\mathbf{Q}(\mathbf{x},\mathbf{p})$ as well.

This because $(\mathbf{x})\mathbf{Q}(\mathbf{x},\mathbf{p})$ (being the proposition expressed by the formula represented by 17 Gen **r**) is the consequent of (2) on the basis of

$$\overline{\mathrm{Bew}_{\mathbf{c}}}(17\ \mathrm{Gen}\ \mathbf{r}) \equiv (\mathbf{x})\overline{\mathbf{x}\ \mathbf{B}_{\mathbf{c}}\mathbf{Sb}\,(\mathbf{p}^{19}_{\mathbf{Z}(\mathbf{p})})} \equiv (\mathbf{x})\mathbf{Q}(\mathbf{x},\mathbf{p});^{3}$$

thus the formula expressing $(\mathbf{x})\mathbf{Q}(\mathbf{x},\mathbf{p})$ and the one expressing $\overline{\mathrm{Bew}_{\mathbf{c}}}(17\ \mathrm{Gen}\ \mathbf{r})$ are equal formulas, associated to the G-number 17 Gen **r**; this being confirmed by the fact that, as the correspondence between G-numbers and formulas is a one-one relation, 17 Gen **r** represents one and the same formula; that is (in Peirce's terminology) all the *tokens* of the same *type* of this formula.

But Gödel's reasoning maintains that **w** represents Wid(**c**) (as stated in (a)), while obliterating that 17 Gen **r** represents $\overline{\mathrm{Bew}_{\mathbf{c}}}(17\ \mathrm{Gen}\ \mathbf{r})$ (as stated in (b)). This will be demonstrated:

first in case that $\overline{\mathrm{Bew}_{\mathbf{c}}}$(17 Gen **r**) is considered in itself, as an arithmetical proposition (see §4),

then in case that $\overline{\mathrm{Bew}_{\mathbf{c}}}$(17 Gen **r**) is taken as representing the correlative metamathematical proposition (see §5).

§4. Gödel says that (**w** Imp(17 Gen **r**) being **c**-*provable*) "if **w** were **c**-*provable*, 17 Gen **r** would also be **c**-*provable*", i.e.

(3) if $\mathrm{Bew}_{\mathbf{c}}$(**w**), then $\mathrm{Bew}_{\mathbf{c}}$(17 Gen **r**),

"and hence it would follow, by (1), that **c** is not consistent",[7] because β) $\mathrm{Bew}_{\mathbf{c}}$(17 Gen **r**), being the consequent of (3), contradicts α) $\overline{\mathrm{Bew}_{\mathbf{c}}}$(17 Gen **r**), being the consequent of (1).

§4a. This conclusion presupposes that it would follow, from (3), that
γ) if c is consistent (or an equivalent, say Wid(**c**)) is provable in c, then $\mathrm{Bew}_{\mathbf{c}}$(17 Gen **r**).

This being confirmed by the fact that Gödel's last conclusion, that "the formula which states that **c** is consistent" (i.e. the one expressing Wid(**c**)) "is not provable in **c**", negates the antecedent of γ,[8] and not only the antecedent of (3).

§4b. But *if*, in (the antecedent of) (3), **w** is considered as representing Wid(**c**), i.e. c is consistent (as presupposed in passing to the antecedent of γ), *then* also 17 Gen **r**, as named by '17 Gen **r**' occurring in (the consequent of) (3), i.e. in β, has to be considered as representing $\overline{\mathrm{Bew}_{\mathbf{c}}}$(17 Gen **r**); *and therefore* the consequent of γ would be not $\mathrm{Bew}_{\mathbf{c}}$(17 Gen **r**), but "$\overline{\mathrm{Bew}_{\mathbf{c}}}$(17 Gen **r**) is provable in c". Thus instead of γ we would have

δ) if c is consistent (or Wid(**c**)) is provable in c, then $\overline{\mathrm{Bew}_{\mathbf{c}}}$(17 Gen **r**) is provable in c.[9]

As the consequent of δ is not contradicting the consequent of (1), on the basis of δ and (1) we cannot draw the known conclusion.

§4c. But Gödel obliterates the feature (just pointed out, see §4b) of the consequent of (3), i.e. of β, exactly when (instead of passing from (3) to δ) he passes from (3) to γ, as presupposed by his conclusion (see §4a). Thus in Gödel's argument:

a) '17 Gen **r**' occurs in '**w** Imp(17 Gen **r**)' in as much as the number 17 Gen **r** represents the formula expressing α) $\overline{\mathrm{Bew}_{\mathbf{c}}}$(17 Gen **r**);

b) '17 Gen **r**' is no more taken in this same way, when occurring in β) $\mathrm{Bew}_{\mathbf{c}}$(17 Gen **r**), though we infer that the proposition expressed by β follows from Wid(**c**) being provable in c (as stated in γ), only on the basis of **w** Imp(17 Gen **r**).[10]

Notice that '17 Gen **r**' occurring in β is no more considered in the above way, (i.e. as stated in (a)), just because the proposition expressed by β is considered as contradicting the one expressed by α. But if, on the other side, we never consider the number 17 Gen r as representing $\overline{\text{Bew}_c}$(17 Gen r), then the reason why 17 Gen r is employed to define w Imp(17 Gen r) fails; and thus no reasoning can be grounded upon w Imp(17 Gen r): neither Gödel's argument (obtaining (3) and γ), nor the one here expounded (obtaining (3) and δ). Thus in Gödel's proof we change the way we take the occurrence of '17 Gen r', passing from 'w Imp(17 Gen r)' to β.[11] Q.e.d.

§4d. If, at the contrary, '17 Gen r' occurs in β as much as the number 17 Gen r represents the formula g expressing $\overline{\text{Bew}_c}$(17 Gen r), then (and only in this case):

1) the proposition expressed by β follows (on the basis of w Imp(17 Gen r)) from Bew_c(w), which represents "Wid(c) is provable in c"; (and in this case we have δ instead of γ: see above, §4b);
but,

2) in this way the proposition expressed by β does not contradict the proposition expressed by α: as '17 Gen r' occurring in α stands *only* for the number 17 Gen r, while '17 Gen r' occurring in β represents, through the number 17 Gen r, the proposition $\overline{\text{Bew}_c}$(17 Gen r). Q.e.d.

§4c. Therefore if words must always be taken (in any argument) to refer to the same entity (although referring to it by any device), Gödel's proof of theorem XI does not hold.

§5. The above argument holds good even if (as usually) we consider 17 Gen r as representing a phrase, i.e. Gödel's undecidable sentence g_u (§3, 2nd case). To expound the argument for this case I shall exhibit the correspondences between G-numbers, formulas, arithmetical and metamathematical propositions here employed. (Column 2 shows the relations and connectives holding between two propositions, formulas, or numbers, indicated in columns 1 and 3; explanations in the beginning of each row refer to columns 1 and 3[12]).

The metamathematical propositions	c is consistent	→	there is no proof, in c, of g_u	
abbreviated	$cns(c)$	→	$\overline{Pf}(g_u)$[13]	(M1)
are represented by the arithmetical propositions	Wid(c)	→	$\overline{\text{Bew}_c}$(17 Gen r)	(1)
equivalent by definition to	Wid(c)	→	(x)Q(x,p)	(2)
being formally expressed by the formulas	w	⊃	g	
associated with the G-numbers	w	Imp	17 Gen r	

On the basis of w Imp (17 Gen r) being c-*provable*, Gödel states (3):

The arithmetical propositions	$\text{Bew}_c(\text{w})$	$\rightarrow$	$\text{Bew}_c(17 \text{ Gen } \mathbf{r})$	(3)
represent the metamathematical ones	$Pf(w)$	$\rightarrow$	$Pf(g)$	(M3)

Notice that g and g_u are two equal formulas. Moreover $\overline{\text{Bew}}_c(17 \text{ Gen } \mathbf{r})$ and $(\mathbf{x})\mathbf{Q}(\mathbf{x},\mathbf{p})$ (being equivalent by definition) are expressed by g.

Column 3 shows the correspondences that are the very basis of the proof, i.e. the following ones. The G-number 17 Gen **r**, as employed to define **w** Imp(17 Gen **r**), represents g; g expresses $(\mathbf{x})\mathbf{Q}(\mathbf{x},\mathbf{p})$ and $\overline{\text{Bew}}_c(17 \text{ Gen } \mathbf{r})$ as well; 17 Gen **r**, being the argument of the proposition $\overline{\text{Bew}}_c(17 \text{ Gen } \mathbf{r})$, is the G-number of a formula g_u, and therefore this proposition (being the consequent of (1)) represents $\overline{Pf}(g_u)$.[13] Moreover $\text{Bew}_c(17 \text{ Gen } \mathbf{r})$, being the consequent of (3), represents $Pf(g)$.[14] The difference between g and g_u is that: g expresses $\overline{\text{Bew}}_c(17 \text{ Gen } \mathbf{r})$, while g_u is represented by 17 Gen **r** being the argument of $\overline{\text{Bew}}_c(17 \text{ Gen } \mathbf{r})$. And this difference is indispensable to the validity of both couples of consequences:

1) (1) and M1: (1) representing M1, and (1) (and (2) as well) being expressed by $w \supset g$ (see §5b);

2) (3) and M3: (3) representing M3, and both being derived from $w \supset g$ (see §5c).

But the proof is valid only if we eliminate this difference (see §5d). This is the clue of the criticism I would advance, concerning the validity of the proof of theorem XI.

§5b. The indispensability of the above difference between g_u and g may even be illustrated as follows.

(1) is stated: 1) because the consistency of c entails $\overline{\text{Bew}}_c(17 \text{ Gen } \mathbf{r})$, which represents $\overline{Pf}(g_u)$; and, 2) in as much as Wid(**c**) represents $cns(c)$.

(2) is derived from (1) because their consequents are equivalent by definition.

Thus the reason why **w** Imp(17 Gen **r**) represents (2) and is **c**-*provable* depends on two facts: 1) 17 Gen **r** (as employed to define **w** Imp(17 Gen **r**)) is the G-number of the formula g, expressing $(\mathbf{x})\mathbf{Q}(\mathbf{x},\mathbf{p})$, i.e. $\overline{\text{Bew}}_c(17 \text{ Gen } \mathbf{r})$, which represents $\overline{Pf}(g_u)$; 2) **w** is the G-number of the formula w, expressing Wid(**c**), which represents $cns(c)$. Thus it may be provable in c that: w implies g (as represented by **w** Imp(17 Gen **r**)), in as much as w expresses Wid(**c**), and g expresses $\overline{\text{Bew}}_c(17 \text{ Gen } \mathbf{r})$; as $\overline{\text{Bew}}_c(17 \text{ Gen } \mathbf{r})$ represents $\overline{Pf}(g_u)$, g cannot be put on a pair with g_u, being argument of $\overline{Pf}(g_u)$. I.e. g and g_u cannot be interpreted, in this context, as expressing one and the same proposition.

§5c. As a consequence of this difference (between g and g_u) we have the following situation. If the correspondences between 17 Gen **r** and g, and between g and $\overline{\text{Bew}}_c(17 \text{ Gen } \mathbf{r})$, would not be eliminated, we could just pass from **w** Imp(17 Gen **r**) to assert that:

if the formula w (represented by $\mathbf{w}$) expressing Wid($\mathbf{c}$) is provable in c, i.e. $Pf(w)$, *then* the formula g (represented by 17 Gen $\mathbf{r}$) expressing $\overline{\mathrm{Bew}}_{\mathbf{c}}$(17 Gen $\mathbf{r}$) is also provable in c, i.e. $Pf(g)$; that is

$$\text{M3)}\quad Pf(w) \rightarrow Pf(g)$$

(M3 being represented by (3)). There is obviously no contradiction between $Pf(g)$, being the consequent of M3, and $\overline{Pf}(g_u)$, being the consequent of M1 (M1 being represented by (1)).

§5d. On the contrary, the conclusion that w is not provable in c could be drawn only if $Pf(g)$ would contradict $\overline{Pf}(g_u)$; but this contradiction (between $Pf(g)$ and $\overline{Pf}(g_u)$) holds only in so far as the difference between g and g_u is obliterated. And this contradiction is the very basis of the proof.

§6. The reason why g and g_u do not express, in this context, the same statement is simply that: g has been introduced here (as implicated by w) only because it is expressing the proposition $\overline{\mathrm{Bew}}_{\mathbf{c}}$(17 Gen $\mathbf{r}$) which represents $\overline{Pf}(g_u)$. Moreover as g expresses the proposition $\overline{\mathrm{Bew}}_{\mathbf{c}}$(17 Gen $\mathbf{r}$) and g_u is represented by the number 17 Gen $\mathbf{r}$ (being the argument of $\overline{\mathrm{Bew}}_{\mathbf{c}}$(17 Gen $\mathbf{r}$)), if we put g on a pair with g_u, then we blurr the difference between: 1) the relation of a formula to the proposition expressed by it, 2) the relation of a formula to the associated G-number. Of course g and g_u have the same (syntactical) form; (and therefore they are represented by the same G-number 17 Gen $\mathbf{r}$). But in the context of this proof g differs from g_u for its meaning; i.e. because w implies g if and only if g expresses $\overline{\mathrm{Bew}}_{\mathbf{c}}$(17 Gen $\mathbf{r}$) which represents $\overline{Pf}(g_u)$. Therefore the difference between g and g_u has to be preserved or not, according to the point of view we assume.

§7. From a logico-philosophical point of view any term, introduced in a demonstration with a given determinate meaning, has to preserve it for the whole proof. Hence in this proof g has always to express $\overline{\mathrm{Bew}}_{\mathbf{c}}$(17 Gen $\mathbf{r}$) and to represent $\overline{Pf}(g_u)$, as it does at the beginning. And therefore, if g_u is taken as expressing or representing any statement, then this same statement can be neither meant nor represented also by g, for the very reason that g represents $\overline{Pf}(g_u)$ (through $\overline{\mathrm{Bew}}_{\mathbf{c}}$(17 Gen $\mathbf{r}$)).

If, from another point of view, the identity of the syntactical form is sufficient (or some peculiar procedure, e.g. the diagonal method, makes it sufficient) to identify them—although their meanings are initially different—then g has to be considered as identical to g_u; thus $Pf(g)$ contradicts $\overline{Pf}(g_u)$, and w is not provable in c (see 5d). Hence, from such a point of view, my critique cannot be stated.

§8. Anyhow the elimination of the distinction between:

1) the formula g_u, represented by the argument of α) $\overline{\mathrm{Bew}}_{\mathbf{c}}$(17 Gen $\mathbf{r}$),
2) the formula g , expressing α and represented by the argument of β) $\mathrm{Bew}_{\mathbf{c}}$(17 Gen $\mathbf{r}$),

is presupposed by the validity of the proof of theorem XI, being based on the contradiction between the propositions expressed by α and β.

NOTES

1. Gödel, Kurt, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme." In *Monatsh. Math. Phys.*, Vol. 38, 1931, pp. 173-198.

2. Braithwaite has shown, on an idea of Rosser, how to construct the formula expressing Wid(**c**). See *R. B. Braithwaite, F.B.A.*, "Introduction to *Gödel, K.*, On Formally Undecidable . . . ", English translation by *B. Meltzer*, Ph.D., Edinburg and London, 1962.

3. On the basis of: 6.1) $\mathrm{Bew}_{\mathbf{c}}(\mathbf{x}) \equiv (E\,\mathbf{y})\,\mathbf{y}\,B_{\mathbf{c}}(\mathbf{x})$, 13) 17 Gen **r** = Sb $(\mathbf{p}^{19}_{Z(\mathbf{p})})$, and 8.1) $Q(\mathbf{x}, \mathbf{y}) \equiv \overline{\mathbf{x}\ B_{\mathbf{c}}\mathrm{Sb}\ (\mathbf{y}^{19}_{Z(\mathbf{y})})}$; ('≡' being used to mean "equivalent by definition"; see Gödel 1931, note 33).

4. I use italics as names of formulas and of classes of formulas of the formal system (see note 12).

5. Pairs of single inverted commas are used to form the name of the word or phrase included; pairs of double inverted commas are used only to single out a phrase, leaving it in formal supposition. Thus, e.g., 'dog' is a word, and "dog" an animal.

6. I consider α and β as names of the phrases written on the same line.

7. I quote from the English translation by *B. Meltzer* (op. cit. at note 2), but the numbering of formulas is changed; i.e. (1) is Gödel's (23), and (2) is Gödel's (24) (see above, § 1).

8. This negation (of the antecedent of γ) obviously depends on the fact that the antecedent of γ is represented by the antecedent of (3).

9. In other words: on the basis of **w** Imp(17 Gen **r**):

   if the formula *w* expressing Wid(**c**) (represented by **w**) is provable in *c*, *then* the formula *g* expressing $\overline{\mathrm{Bew}_{\mathbf{c}}}$(17 Gen **r**), and represented by 17 Gen **r** as named by the occurrence of '17 Gen **r**' in β, is provable in *c*;

   and therefore: if the proposition $\mathrm{Bew}_{\mathbf{c}}(\mathbf{w})$ represents the one saying that "the formula *w*, expressing Wid(**c**), is provable in *c*"; *then* the proposition $\mathrm{Bew}_{\mathbf{c}}$(17 Gen **r**) (following from $\mathrm{Bew}_{\mathbf{c}}(\mathbf{w})$, and expressed by β) represents the one saying that "the formula *g*, expressing $\overline{\mathrm{Bew}_{\mathbf{c}}}$(17 Gen **r**), is provable in *c*".

   Thus from (3) we obtain δ (instead of γ).

10. That is to say: when on the basis of **w** Imp(17 Gen **r**) we pass to affirm that "if **c** is consistent is **c**-*provable*, then 17 Gen **r** is **c**-*provable*", thus demonstrating that "**c** is consistent is not **c**-*provable*", we simply "forget" the reason why 17 Gen **r** is employed to define **w** Imp(17 Gen **r**). I.e. we fail to remember that 17 Gen **r** has here been introduced as G-number of the formula *g*, which expresses $\overline{\mathrm{Bew}_{\mathbf{c}}}$(17 Gen **r**); though we still maintain, along the whole proof, that **w** represents Wid(**c**), being equivalent to "*c* is consistent".

11. In fact '17 Gen r' occurs in 'w Imp(17 Gen r)' in as much as the number 17 Gen r represents the formula expressing $\overline{\mathrm{Bew}_c}$(17 Gen r); and at the end of the argument, i.e. in β, Gödel considers this occurrence of '17 Gen r' as naming only 17 Gen r, which does no longer represent $\overline{\mathrm{Bew}_c}$(17 Gen r).

12. I use italics for typographical abbreviations of expressions for metamathematical concepts and for names of formulas and of classes of formulas ($\supset$ being in autonomous use); and Hilbert's notation (following Gödel).

    Thus c and g_u are names of the class of formulas and formula represented, respectively, by c and by 17 Gen r, as occurring in (1); g is the name of the formula expressing $\overline{\mathrm{Bew}_c}$(17 Gen r), equivalent to (x)Q(x, p) (see, respectively, (1) and (2)); and g is represented by 17 Gen r, as employed to define w Imp(17 Gen r).

13. As in this case we consider 17 Gen r as representing the formula g_u, we shall also consider the arithmetical predicate Bew_c(x) as representing the metamathematical one *Pf*(x).

14. The reason is that (3) depends from w Imp(17 Gen r), and that 17 Gen r is employed to define w Imp(17 Gen r) because it is the G-number of g.

Università degli' studi di Perugia
Perugia, Italy