A Critique of Self-Reference: What Gödel's Theorem Really Proves

Antonio Leon
Instituto F. Salinas (Retired), Salamanca, Spain.

Abstract.-This article introduces a new perspective for the analysis of self-referential sentences, and proves the conditions under which they are inconsistent. The Liar Paradox, Grelling-Nelson Paradox, Russell's Predicate Paradox, Russell's Set Paradox and Richard Paradox are proved to meet such conditions. The same is proved of the ordinary language interpretation of Gödel's undecidable formula if the corresponding formal calculus is complete. In consequence, Gödel's Theorem VI only holds if that calculus is not complete, which makes the theorem unnecessary. All proofs and arguments in this article are developed within the framework of a simplified system of ordinary logic also defined in this paper.

Keywords: self-reference, self-subject, theorem of the inconsistent self-subjects, liar paradox, Grelling-Nelson paradox, Russell's predicate paradox, Russell's set paradox, Richard paradox, Gödel's first incompleteness theorem.

1 Introduction

Gödel's First Incompleteness Theorem (Theorem VI in his seminal paper of 1931) was deduced within a formal calculus C (P in Gödel's paper) of which Gödel wrote [4, p. 176]:

> By and large, C is the system that you get by building the logic of [Russell-Whitehead Principia Mathematica] PM on top the Peano axioms (numbers as individuals, successor-relation as undefined basic concept).

In order to express in arithmetical terms the formal relations between the abstract chains of symbols of his formal calculus C, Gödel introduced in the first part of his 1931 article, a precise codification of all elements of C based on the Fundamental Theorem of Arithmetics, according to which each natural number has an exclusive and unique decomposition into prime factors. This arithmetical fact makes it possible to codify each element of C: constants signs, numerical variables, predicate variables, formulas, sequences of formulas (demonstrations) etc., even certain meta-mathematical sentences as: this sequence of formulas is a demonstration of this or that formula; this formula contains this or that variable; etc. Gödel then defines a formula G, a chain of C-symbols, that once codified would be a gigantic natural number which, fortunately, is not necessary to calculate. But if it were calculated it could be precisely decoded as a specific and unique chain of C-symbols. These symbols are meaningless in the sense that C-theorems are deduced exclusively from their syntactic rules of inference without attending symbols meaning. Notwithstanding, C-rules of inference were devised so that it were possible to interpret its symbols in ordinary language by observing their formal behaviour. For example, in the three chains of symbols:

$$0+s(0)=s(0); \qquad s(0)+s(0)=ss(0); \qquad 0+s(0)+s(0)=ss(0) \tag{1}$$

the symbols 0, +, s, (,), and = can be respectively interpreted as *zero*; *plus*, *successor of*, *left parenthesis*, *right parenthesis*, *equal*. So, Gödel's formula G is a chain of C-symbols that are unequivocally interpreted in ordinary language as a self-referential meta-mathematical sentence that assert of itself that it is not demonstrable in C (C-dem hereafter):

$$G: \ G \text{ is not C-dem} \tag{2}$$

For the same reason the C-chain of symbols $0 + 0 = 0$ is interpreted in ordinary language as: *zero plus zero equal zero*, Gödel's formula G is interpreted in ordinary language terms as (2). It will not be necessary to go into further details on Gödel's complex formalism because all arguments and proofs in this paper will be self-sufficient and developed within the framework of ordinary logic, the familiar logic subsumed into natural languages (a brief evaluation of ordinary logic is given in the next section).

Once defined his formula G, Gödel proved in Theorem VI that if G were C-dem, then $(\neg G)$ would also be C-dem, which is impossible if C is consistent. And if $(\neg G)$ were C-dem, then C would be inconsistent (really only ω-inconsistent, see Appendix on ω-inconsistency). So, G is undecidable and true. In consequence, C is not complete. On this kind of arguments, and in the same 1931 paper, Gödel wrote [4, p. 175-176]:

> The analogy of this conclusion with the Richard-antinomy leaps to the eye; there is also a close kinship with the liar-antinomy, [...]

For this reason this paper focuses its attention on self-referential paradoxes, albeit from a completely new perspective: the perspective of the self-subjects introduced later in this work.

2 Nature, Language and Logic

Gödel's formula G is an abstract meta-mathematical formula whose veracity can be tested once decoded and interpreted in terms of ordinary language. It is in this way that the first order logic of an abstract formal calculus as C and the ordinary logic of natural languages meet each other. The encounter will be analyzed in Section 7, and within the framework of the ordinary logic system introduced in the next section. The objective of this short section is to justify the use of ordinary logic for such a purpose. By way of preamble, let us recall that ordinary logic is an attribute of rational life, the consequence of a physical world evolving under a consistent set of laws.

In effect, at the micro and mesoscopic scales where organisms evolve, nature appears to be clearly consistent, subjected to a set of invariable rules (at least for periods of time sufficiently large): the laws of physics. No exception is known to this general conclusion. Living beings evolve and thrive in syntony with the logical consistency of nature. They have been forced, in evolutionary terms, to inscribe the logic of nature in their own nature: they need to know the way nature works to perform the appropriate responses in order to survive and reproduce inside it [6, p. 357-361]. It is that inevitable dependence that force organisms to evolve as consistent systems modeled by natural selection in agreement with natural logic. Life reflects the formal consistency of the universe and makes our ordinary logic a safe formal instrument to get conclusions on what happens, and previsions on what will happen, in the physical world. In short, ordinary logic is endorsed by the rigorous formal consistency of nature. In this sense, let us recall that formal sciences, experimental sciences, human sciences, first (and superior) order logic, and ordinary logic share, all of them, at least the two first of the following three laws (p and q stand for propositional sentences):

$$p \Rightarrow p \quad \text{First Law (Law of Identity, see below).} \tag{3}$$

$$\neg(p \wedge \neg p) \quad \text{Second Law (Law of Non-Contradiction).} \tag{4}$$

$$p \vee \neg p \quad \text{Third Law (Law of the Excluded Middle).} \tag{5}$$

These laws reflect the logical essence of the universe and ordinary logic. Thus, ordinary logic is not arbitrary but grounded on the logic of nature. Ordinary logic, in turn, inspired and promoted the birth of formal logic, the birth of logic as a scientific discipline. Formal logic improves ordinary logic by concentrating on form up to the point of becoming independent of the meaning that characterizes ordinary language. Notwithstanding, abstract formulas of formal calculi can be interpreted in terms of ordinary language, what makes it possible to check their veracity and consistency from the perspective of ordinary logic and empirical testing.

3 The Ordinary Logic System

This section introduces the formal system that will be used in the subsequent discussions. It is a very simplified formal system that will be referred to as ordinary logic system (OLS). It is composed of the following elements:

a) Sentences: declarative sentences of ordinary language each assigned with an unequivocal and exclusive meaning (proposition), and being each of them composed of only one singular subject; only one verb: either *is* or *is not*; and only one singular predicate.

b) Names of sentences: symbols, here lowercase letters, that stand for the sentences, and that can replace anywhere the corresponding sentences.

c) The logic symbols $\neg$ (negation); $\Rightarrow$, (implication); = (equality).

d) Syntactic rules: syntactic rules of ordinary language, including the rule of double negation, that state the way of forming well formed sentences.

e) A set of axioms: the three laws of logic (3)-(4)-(5).

f) A set of inference rules to get conclusions from different sentences:

   (a) Modus ponens: $((p \Rightarrow q) \wedge p) \Rightarrow q$

   (b) Modus tollens: : $((p \Rightarrow q) \wedge \neg q) \Rightarrow \neg p$

   (c) Hypothetical syllogism: $p \Rightarrow q \Rightarrow r \Rightarrow s \Rightarrow \cdots \Rightarrow z$. Therefore $p \Rightarrow z$

were $p, q, r, \ldots$ are names of sentences. Self-referential sentences will be written in the canonical form introduced above (2):

$$p\text{: } p \text{ is } X \tag{6}$$

OLS only considers well formed sentences. For example, if p is the name of the sentence [S is X], the following sentences are all of them well formed sentences:

$$p: \quad S \text{ is X} \tag{7}$$

$$\neg p: \neg(S \text{ is } X) = S \text{ is not } X \tag{8}$$

$$\neg\neg p: \neg(\neg(S \text{ is } X)) = \neg(S \text{ is not } X) = S \text{ is } X \tag{9}$$

Evidently, true sentences cannot be inconsistent. Otherwise a truth would imply a falsehood through one of the terms of the involved contradiction. In OLS it is assumed that a sentence can take the following values:

a) True: the case is what the sentence asserts to be.

b) Not true (false): the case is not what the sentence asserts to be.

c) Undecidable: true or not true but impossible to decide which is the case.

d) Inconsistent: it implies a contradiction, i.e a sentence and its negation.

Since the subject of a self-referential sentence is itself a sentence, it can take the same above values a)-d). For this reason, the subjects of self-referential sentences will be called *self-subjects*. Accordingly, they can be true, false, undecidable or inconsistent.

4 The First Law of Logic

The First Law of logic is usually stated in Aristotelian terms such as: *A thing is what it is, and it is not what it is not*. Or as $A = A$. From the point of view of (propositions) sentences, the First Law can be written [1, p. 139] as:

$$p \;\Rightarrow\; p \tag{10}$$

where p is any sentence. Implication (10) translates the sense of identity to the world of sentences. Notice that, in fact, implication (10) is always true, be p true or not; (10) it is a tautology that must hold in all formal calculi. It is immediate, on the other hand, to deduce the Aristotelian formulation from (10). Indeed, assume $A \neq A$; we would have two different instances of A, say A and A'; and being different, one of them, for instance A, could be false and the other true. Therefore, the implication $A \Rightarrow A'$ would be false. So, it must be $A = A$.

5 Theorem of the Inconsistent Self-Subjects

The next theorem states the conditions under which self-subjects are inconsistent.

Theorem of the Inconsistent Self-Subjects.-*If S is the self-subject of a predicate X, and S and X satisfy:*

$$S \textit{ is } X \Rightarrow S \textit{ is not } X. \tag{11}$$

$$S \textit{ is not } X \Rightarrow S \textit{ is } X. \tag{12}$$

then S is an inconsistent self-subject of the predicate X, and the sentences [S is X] and [S is not X] are both inconsistent.

Proof.-Consider the sentences p and q:

$$p : S \text{ is } X. \tag{13}$$

$$q : S \text{ is not } X. \tag{14}$$

Taking into account (7)-(9), we can write:

$$q = S \text{ is not } X. \tag{15}$$

$$q = \neg(S \text{ is } X). \tag{16}$$

$$q = \neg p \tag{17}$$

$$\neg\neg p = \neg q \tag{18}$$

$$p = \neg q \tag{19}$$

On the other hand, (11) and (12) can be written respectively as:

$$p \Rightarrow q \tag{20}$$

$$q \Rightarrow p \tag{21}$$

With respect to p, and according to the First Law of Logic, we have:

$$p \Rightarrow p \tag{22}$$

which in accord with (19), can also be written as:

$$p \Rightarrow \neg q \tag{23}$$

Thus, by (20) and (23), we have:

$$\begin{cases} p \Rightarrow q \\ p \Rightarrow \neg q \end{cases} \tag{24}$$

Therefore, p is inconsistent. Similarly, with respect to q and according again to the First Law, we have:

$$q \Rightarrow q \tag{25}$$

which in accord with (17), can also be written as:

$$q \Rightarrow \neg p \tag{26}$$

Thus, by (21) and (26), we have:

$$\begin{cases} q \Rightarrow p \\ q \Rightarrow \neg p \end{cases} \tag{27}$$

Therefore, q is also inconsistent. Once proved that p and q are both inconsistent, let us analyze the alternatives for the self-subject S:

a) If S were X then p would be true, what is impossible because p is inconsistent.

b) If S were $\neg X$ then q would be true, what is impossible because q is inconsistent.

c) If S were undecidable either p or q would be true (though we could not decide which is the case), what is impossible because both of them are inconsistent.

Therefore, S can only be inconsistent. And taking into account the contradictions (24) and (27), the sentences [S is X] and [S is not X] are inconsistent. □

6 Consequences on some self-referential paradoxes

We will now examine the consequences of the above theorem on five self-referential paradoxes.

- The Liar Paradox:

$$p : p \text{ is not true} \tag{28}$$

On the one hand we have:

$$p \text{ is true} \Rightarrow p \text{ is not true} \tag{29}$$

Because if p is true, then it is true what it affirms; and being what it affirms that p is not true, then p is not true. On the other hand:

$$p \text{ is no true} \Rightarrow p \text{ is true} \tag{30}$$

Because if p is not true then it is not true what p affirms, and being what p affirms that p is not true, then it is not true that p is not true. So, p is true. Therefore, and in accord with the Theorem of the Inconsistent Self-Subjects, the sentence p is an inconsistent self-subject of the predicate *true*, and the sentences [p is true] and [p is not true] are both inconsistent.

- Grelling-Nelson Paradox:

$$\text{Heterologic is not heterologic} \tag{31}$$

where heterologic (H for short) is an adjective that does not describe itself, being not heterologic if it describes itself. On the one hand we have:

$$H \text{ is } H \Rightarrow H \text{ is not } H. \tag{32}$$

because if H is H, then H is an adjective that does not describe itself. Therefore, H is not H. On the other hand we also have:

$$H \text{ is not } H \Rightarrow H \text{ is } H. \tag{33}$$

because if H is not H, then H is not an adjective that does not describe itself. So, it is an adjective that describes itself. Hence, H is H. Therefore, and in accord with the Theorem of the Inconsistent Self-Subjects, *heterologic* is an inconsistent self-subject of the predicate *heterologic*, and the sentences [H is H] and [H is not H] and are both inconsistent.

- Russell's Predicate Paradox [2]:

$$\text{Russellian is not Russellian.} \tag{34}$$

where Russellian (Rs for short) is a predicate that does not predicate itself. And not Russellian a predicate that predicates itself. On the one hand we have:

$$Rs \text{ is } Rs \Rightarrow Rs \text{ is not } Rs. \tag{35}$$

because if Rs is Rs, then Rs does not predicate itself. So, Rs is not Rs. On the other hand:

$$Rs \text{ is not } Rs \Rightarrow Rs \text{ is } Rs. \tag{36}$$

because if Rs is not Rs, then Rs is not a predicate that does not predicate itself. So, it is a predicate that predicates itself. Hence Rs is Rs. Therefore, in accord with the Theorem of the Inconsistent Self-Subjects, *Russellian* is an inconsistent self-subject of the predicate *Russellian*, and the sentences [Rs is Rs] and [Rs is not Rs] are both inconsistent.

- Russell's Set Paradox: the set R all of sets that do not belong to themselves was legitimated by the Axiom of Comprehension: given any predicate, there exists the set of all the elements that satisfy that predicate. As is well known, this set caused a major crisis in the foundation of mathematics. But it is also a self-referential set that falls under the jurisdiction of the Theorem of the Inconsistent Self-Subjects. Indeed, we have:

$$R \text{ belongs to } R \Rightarrow R \text{ does not belong to } R \tag{37}$$

because if R belongs to R, then it is an element of the set whose elements do not belong to themselves, then it does not belong to R. And, on the other hand:

$$R \text{ does not belongs to } R \Rightarrow R \text{ belongs to } R \tag{38}$$

because if R does not belong to R, then R is an element of the set of all sets that do not belong tho themselves, which is just the set R. So, it belongs to R. Therefore R is an inconsistent self-subject of the predicate *to belong to itself*, and the sentences [R belongs to itself], and [R does not belong to itself] are both inconsistent.

- Richard's paradox. Suppose we write a list of all arithmetic properties of the natural numbers, for example to be even; to be odd; to be prime; to be a multiple of 3; and so on. Let us suppose now that we sort the list with some criteria. For example according to the number of letters used in the description of the property, and in alphabetical order for those with the same number of letters. We will have a list with a first element, a second element, a third element etc. That is, an ordered list whose successive elements can be indexed by successive natural numbers. Let us call R_1, R_2, $R_3 \ldots$ the successive element of that list. Suppose that R_{17} corresponds to the property *to be even*, and that R_{1125} corresponds to the property *to be odd*, we will say the index 17 (of R_{17}) is Richardian because the property it indexes cannot be applied to itself: 17 is not even. On the other hand the index 1125 (of R_{1125}) is not Richardian because it can apply to itself the property it indexes: 1125 is odd. In general we will have:

$$R_n = \text{ Property } X \tag{39}$$

and the index (natural number) n will be Richardian if it does not satisfy the property X, and not Richardian if it does.

Initially, it was not taken into account that being Richardian is not an arithmetical property. Whether or not a number is Richardian depends on the arbitrary order we impose on the list of arithmetic properties. We can order that list in many ways, and in each ordering the numbers that are Richardian and those that are not will change. To be Richardian is a metaarithmetical sentence (a sentence on arithmetical properties). So being Richardian should not appear in the list of arithmetic properties of the natural numbers. This problem was solved in Gödel's formal calculus in which it is possible to define formulas and meta-formulas in the same numerical terms. But for our purpose, assume as Richard did, that being Richardian is one of the indexed arithmetic properties, and that it is indexed by the natural number k:

$$k: \text{ to be Richardian} \tag{40}$$

On the one hand we have:

$$k \text{ is Richardian } \Rightarrow k \text{ is not Richardian} \tag{41}$$

because if k is Richardian, then k does not satisfy the property it indexes, which is the property of being Richardian; therefore k is not Richardian. On the other hand:

$$k \text{ is not Richardian} \Rightarrow k \text{ is Richardian} \tag{42}$$

because if k is not Richardian, then k is not a number that does not satisfy the property it indexes. So, k satisfies the property it indexes, which is that of being Richardian. Hence, k is Richardian. Therefore, in accord with the Theorem of the Inconsistent Self-Subjects, k is an inconsistent self-subject of the predicate *Richardian* and the sentences [k is Richardian] and [k is not Richardian] are both inconsistent.

7 Gödel Undecidable Sentence

This section analyzes the case of the ordinary language interpretation of Gödel formula G under the provisional hypothesis that Gödel formal calculus C is a complete calculus. Indeed, assume, *only for a moment*, that Gödel's formal calculus C is complete. We could write:

$$G \text{ is C-dem} \Rightarrow (\neg G) \text{ is C-dem} \qquad \text{(Proved by Gödel)} \tag{43}$$

$$(\neg G) \text{ is C-dem} \Rightarrow G \text{ is not C-dem} \qquad \text{(By C-consistency)} \tag{44}$$

So we can write:

$$G \text{ is C-dem} \Rightarrow G \text{ is not C-dem} \qquad \text{(Hypothetical syllogism)} \tag{45}$$

On the other hand:

$$G \text{ is not C-dem} \Rightarrow (\neg G) \text{ is C-dem} \qquad \text{(By completeness)} \tag{46}$$

$$(\neg G) \text{ is C-dem} \Rightarrow C \text{ is inconsistent} \qquad \text{(Proved by Gödel-Rosser)} \tag{47}$$

$$C \text{ is inconsistent} \Rightarrow G \text{ is C-dem} \qquad \text{(By C inconsistency)} \tag{48}$$

So we can write:

$$G \text{ is not C-dem} \Rightarrow G \text{ is C-dem} \qquad \text{(Hypothetical syllogism)} \tag{49}$$

Therefore, and according to the Theorem of the Inconsistent Self-Subjects, if C were complete, G would be an inconsistent self-subject of the predicate C-dem, and the sentences [G is C-dem] and [G

is not C-dem] would be inconsistent. Therefore, Gödel's Theorem VI is significant if we previously assume C is not complete, in which case the theorem becomes unnecessary. The alternative is that there is no ordinary language interpretation G of G, in which case nothing can be said on G, except that it is a well defined chain of C-symbols. Note that according to (45) and (49) the sentence [G is C-dem] behaves as the above self-referential paradoxes. It could be called Gödel's Paradox.

8 Short epilog

I agree with Galileo's opinion on the Liar Paradox [3, pp 93-94] (italic is mine):

> This is one of those dilemmas called sorites. Like the Cretan, who said that all Cretans were liars, therefore, since he was a Cretan, he was lying when he said that Cretans were liars. Therefore, it was necessary for the Cretans to be truthful and, consequently, he was truthful and, therefore, in saying that the Cretans were liars, he was telling the truth, and being one of the Cretans he had to be a liar. *And so, in this kind of sophistry, one would be going round and round eternally without ever concluding anything.*

And also with Wittgenstein [8, 3.332, 3.333]:

> **3.332**. No proposition can state something about itself, since the propositional sign cannot be contained in itself (this is the whole of Type Theory).
> **3.333**. A function cannot be its own argument, because the functional sign already contains the prototype of its own argument and it cannot contain itself.

In my opinion self-referential sentences and formulas are capricious games of words and symbols that have made us waste a lot of time and money. Until now they have been absolutely useless in order to explain the physical world. Recall, on the other hand, that self-reference had to be axiomatically removed (Axiom of Regularity) from set theory to avoid certain persistent inconsistencies. I think that ordinary language and formal languages should follow the example of set theory with respect to self-reference.

Appendix on ω-inconsistency

Gödel originally proved that if $\neg G$ is C-dem then C is ω-inconsistent. A formal calculus, as Gödel's calculus C, is ω-inconsistent, if for some arithmetic predicate P it is possible to prove the following two results:

(A): There is at least one number n that satisfies the predicate P: $(\exists n)P(n)$

(B): The predicate P is not satisfied for the successive natural numbers: $\neg P(1)$, $\neg P(2)$, $\neg P(3), \dots$

As Gödel himself pointed out, an inconsistent system is also ω-inconsistent, but the converse is not true: a system can be ω-inconsistent and consistent [5, p. 72]. In 1936, B. Rosser proved for another self-referent formula G', similar to Gödel's G, that if G' is C-dem then C is inconsistent [7]. It can be proved, however, that Rosser's improvement is unnecessary because it is possible to prove that, contrarily to what is assumed since Gödel's time, ω-inconsistent systems are also inconsistent. The proof makes use of a supertask. Note a supertask is not a proof but a sequence of tests, all of which can be carried out in a finite interval of time according to the hypothesis of the actual infinity subsumed into the Axiom of infinity, so that if we assume the axiom we must also assume supertasks.

Consider the following supertask S: Define the boolean variable b with the initial value $b = false$, and let $\langle t_i \rangle$ be an ω-ordered sequence of instants within the finite real interval (t_a, t_b) whose limit is t_b. At each of the successive instants t_1, t_2, $t_3, \dots$ of the sequence $\langle t_i \rangle$ test if the successive natural numbers 1, 2, 3,... satisfy the predicate P; each number $n \in \mathbb{N}$, checked at the instant $t_n \in \langle t_i \rangle$. If n satisfies P, and only if it satisfies P, the variable b is redefined as *true* and the supertask ends. If n does not satisfy the predicate P, the following number $n+1$ is checked. Being t_b the limit of the sequence $\langle t_i \rangle$, the one to one correspondence f between $\mathbb{N}$ and $\langle t_i \rangle$ defined by $f(n) = t_n$, $\forall n \in \mathbb{N}$, proves that, in any case, at instant t_b the supertask S has already finished. Let us examine the two alternatives for the value of b at instant t_b, which are mutually exclusive and exhaustive.

1) b = false: no natural number k exists such that $(\exists n)P(n)$. So (A) is false and cannot be proved.
2) b = true: A natural number k has been found such that $P(k)$. So (B) is false and cannot be proved.

Therefore, it is impossible to prove both (A) and (B). In consequence, ω-inconsistency is inconsistent. So, ω-inconsistent calculi are also inconsistent.

References

[1] Calixto Badesa, Ignacio Jané, and Ramón Jansana, *Elementos de lógica formal*, Editorial Ariel (col. Ariel Filosofía), Barcelona, 2000.

[2] John W. Dawson, *Logical Dilemmas. The life and work of Kurt Gödel*, A K Peters Ltd., Wellesley MA, 1997.

[3] Galileo Galilei, *Diálogo sobre los dos máximos sistemas del mundo ptolemaico y copernicano*, Círculo de Lectores, Barcelona, 1997.

[4] Kurt Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. I*, Monatshefte für Mathematik und Physik **38** (1931), 173–198.

[5] Kurt Gödel, *Obras completas*, Alianza, Madrid, 1989.

[6] Antonio Leon, *Infinity put to the test*, Printed at Amazon, 2021.`https://www.amazon.com/dp/B08WP3L21T`

[7] Barkley Rosser, *Extensions of some theorems of gödel and church*, Journal of Symbolic Logic **1** (1936), no. 3, 87–91.

[8] Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, Alianza, Madrid, 1999.