THE REFLEXIVITY OF A LANGUAGE (SOME COMMENTS ON GÖDEL AND WITTGENSTEIN)

Virgil DRĂGHICI[1]

[1]Assoc. Prof. PhD, Babes-Bolyai University Cluj-Napoca, Romania
Corresponding author: Virgil Drăghici; e-mail:vidra007@gmail.com

Abstract

In its special forms of self-reference or diagonalization, the reflexivity of language leads to the paradoxes. How do we manage these paradoxical constructions? In order to answer the question I take two paradigmatic cases: Gödel and Wittgenstein. Even if in both cases the starting point is the same, the Russell's Paradox, as section 2 of this paper shows, the managements of these constructions are very different. In the first case we speak about fundamental results of mathematical logic with substantial philosophical applications: the limitative theorems. Section 1 exemplifies how to construct a Gödel-type sentence using the Diagonal Lemma, and proves its undecidability in Peano Arithmetic. Section 3 argues how the diagonalization leads, *via* Kleene's Theorem, to the Turing form of Gödel's Theorem, the core of Lucas/Penrose Argument. In the second case, instead, as section 4 shows, the same procedure of diagonalization suppresses the reflexivity of a language in Wittgenstein's *Tractatus*, leading finally to the self-suppressing of the philosophical discourse.

Keywords: *reflexivity, self-reference, diagonalization, paradox, Gödel, Wittgenstein.*

1. PRELIMINARY

As Tarski rightly remarked, the natural languages are semantically closed, i.e. they contain semantic terms (true, denotation) and have the possibility of referring not only to the objects, but also to their own expressions. These facts, *via* classical logic, open the possibility of constructing self-referring expressions and *eo ipso* to generate semantic paradoxes. Let us call this property of a language, to speak about itself, *reflexivity* (TARSKI, 1936; TARSKI, 1944; TARSKI, 1956).

Are these paradoxical constructions relevant in any way? As it is well-known, they are, both in a logical/mathematical and in a philosophical sense. Let us give an example. A Liar-type sentence is a sentence asserting its own falsity, like "This sentence is not true." If in this sentence we replace the semantic term "true" with the syntactic term "provable," we obtain a sentence of Gödel-type, asserting its own unprovability; let it be G: "This sentence is not provable". Now, let us consider a formal system S and that S is correct (or sound), i.e. it does not prove false sentences. Let us ask the following questions: Is G provable in S? Is it true or false? As can be argued, if G is provable, then it is false, since it asserts its own unprovability, and then it is not provable in S, since S is correct. Hence, by *reductio*, G is not provable in S. But not being provable in S is what G itself asserts, and then G is true. On the other hand, since G is true, it follows that its negation, not-G, is false, and then not provable in S, again by correctness of S. We have therefore a sentence G true but not decidable in S. What exemplifies this heuristic argument is that "true" and "provable" do not coincide. It shows, *in nuce*, a fundamental result of mathematical logic: the incompleteness phenomenon.

2. GÖDEL INCOMPLETENESS THEOREM

The discovery of incompleteness is due to Gödel. It can be obtained by transposing the heuristic argument in a *formal* shape (GÖDEL, 1931; GÖDEL, 1986). Consider in what follows as S the formal system of Peano Arithmetic: PA. Of course, there are many ways to formally construct a G-type sentence (that is, asserting its own unprovability). It is more convenient to take the primitive recursive relation $Pf(y,x)$: "y is (the Gödel number of) a proof of the sentence (with Gödel number) x", and $\Pi(y,x)$ be the formula expressing it formally in PA. Now, let β be the formula $\forall y\neg\Pi(y,x)$, with x free. By *Diagonal Lemma* (DL) there is a sentence G such that

$$PA \vdash G \equiv \forall y\neg\Pi(y,\boldsymbol{g}),$$

where g is the Gödel number of G (BOOLOS et al., 2002).

Gödel's Theorem (for PA)

(a) *If PA is consistent, then G is true but not provable in PA; (b) If PA is ω-consistent, then ¬G is not provable in PA.*

Proof (a)(*Reductio*). Assume hypothesis and suppose that G is provable, i.e. $\vdash G$, equivalently $\vdash \forall y\neg\Pi(y,\boldsymbol{g})$. Let *k* be the Gödel number of a proof of G in PA. Then *Pf*(*k*,*g*) holds, and therefore $\vdash\Pi(\boldsymbol{k},\boldsymbol{g})$ (by formal expressibility of *Pf* in PA). On the other hand, from $\vdash \forall y\neg\Pi(y,\boldsymbol{g})$ it follows, by first order-logic, that $\vdash\neg\Pi(\boldsymbol{k},\boldsymbol{g})$, contradicting the hypothesis of consistency of PA.

(b)(*Reductio*). Assume hypothesis and suppose that $\vdash\neg G$, equivalently, $\vdash \$y\Pi(y,\boldsymbol{g})$. Since by (a) $\nvdash G$, it follows that *Pf* (*y*,*g*) is false for any *y*. Therefore, for any *y*, $\vdash\neg\Pi(\boldsymbol{y},\boldsymbol{g})$. But from $\vdash \$y\Pi(y,\boldsymbol{g})$ and $\vdash\neg\Pi(\boldsymbol{y},\boldsymbol{g})$ for any *y* the ω-inconsistency of PA follows.

The sentence G is called the fixed point of the formula β.

How do we argue that G is true? A simple way is just the proof of (a): supposing that G is false, i.e. G were provable, a contradiction follows. Hence, under the assumption of the consistency of PA, G is true. Another way is the following: G is what is usually called a Π_1-closed formula (sentence), i.e. a sentence of the form $\forall x a(x)$ with a(x) a decidable formula of Lpa (BOOLOS, 1993). As can be proved, the following equivalence holds:

(*Eq1*) PA is Π_1-correct iff PA is consistent.

And then the part (a) of Gödel's Theorem can be converted into (a*) If PA is Π_1-correct (that is, correct for Π_1-sentences), then G is true and not provable in PA. The connection with heuristic argument is straightforward.

Similarly, a Σ_1-sentence is a sentence of the form \$xa(x), with a(x) decidable. For this kind of sentences the following equivalence holds:

(*Eq2*) PA is Σ_1-correct if PA is ω-consistent (SMORYNSKI, 1977).

Can the argument for the truth of G be reproduced *within* PA? The answer is no, according to Tarski's Theorem. Let us give a simple argument for this limitative theorem. Suppose Tr(x) is a formula of Lpa expressing the truth predicate in PA. Let ¬Tr(x) be its negation. By DL there is a sentence S such that $PA \vdash S \equiv \neg Tr(\ulcorner S \urcorner)$. But since Tr(x) is the *truth* predicate for PA, it must obey the Tarski's Schema (T):

$S \equiv Tr(\ulcorner S \urcorner)$, and by the two equivalences we obtain $PA \vdash Tr(\ulcorner S \urcorner) \equiv \neg Tr(\ulcorner S \urcorner)$, and this is the formal shape of a Liar-type sentence. Hence PA would be inconsistent (TARSKI, 1936; TARSKI, 1944).

What do these results show? Escaping from paradox has a cost: the so-called *limitative theorems*: G is not provable (if PA is consistent), ¬G is not provable (if PA is ω-consistent) and the truth predicate for PA is not a formula of Lpa (of its object-language) but of its metalanguage. As it can be seen, all these fundamental results are directly connected to paradoxes. In the next section we take a closer look at the relationship between the paradoxes and the incompleteness phenomenon (GÖDEL, 1931; GÖDEL, 1986).

3. PARADOXES AND THE INCOMPLETENESS PHENOMENON

The Liar Paradox is not the only way to derive the incompleteness results. A more interesting case is to consider two other paradoxes: *Russell's Paradox* and *Grelling Paradox.*

Russell's Theory of Types was invented in order to avoid set-theoretical paradoxes. Such a paradox was constructed in the context of showing that Frege's naïve theory of sets is inconsistent. This theory is based on extensionality and comprehension schema (*Compr*). According to *Compr* every predicate (condition) can define a set or, equivalently, a set is just the extension of an arbitrary predicate, i.e.

$$Compr. \quad (Ey)(x)\ (x \in y \leftrightarrow a(x)),$$

with a(*x*) not containing *y* free. Now, if a(*x*) is the predicate ~(*x*∈*x*), then by *Compr*, (*Ey*)(*x*)

$(x \in y \leftrightarrow \sim(x \in x))$, whence for $x=y$ (diagonalization!) we obtain $(Ey)\ (y \in y \leftrightarrow\ \sim(y \in y))$, and this is what is called Russell's Paradox. What this paradox shows is that unrestricted *Compr* does generate an inconsistency *via* extensionality. Grelling Paradox can be obtained from Russell's Paradox by interpreting "$\in$" (is a member of) as "true of" (*Trof*) and taking for $a(x)$ in *Compr* the predicate $\sim$*Trof* (x,x) ("x is not true of itself"). Whence for $x=y$ we get the following form of Grelling Paradox (a.k.a. Heterological Paradox):

Gr. $\quad (Ey)(\ Trof\ (y,y) \leftrightarrow \sim Trof\ (y,y)$,

That is, "y is true of itself" (autological) if and only if "y is not true of itself" (heterological).

Moreover, if y is the Gödel number of a formula $\alpha(z)$, with z free, and *Sat* (y,x) is the predicate "x does satisfy y", then we have: *Trof* (y,x) iff *Sat* (y,x).

Now, what is the relationship between Gödel's result and Grelling Paradox? On the one hand, as we saw above in a heuristic form, Gödel's result can be obtained from Liar Paradox. And this paradox is a Liar-type sentence, one asserting its own falsity. Hence, a Liar-type sentence will also be the following expression: "'Yields a falsehood when appended to its own quotation' yields a falsehood when appended to its own quotation," a sentence asserting its own falsity. In this way, the connection with Grelling Paradox becomes evident, since a predicate is heterological if it is not true of itself or, equivalently, "yields a falsehood when appended to its own quotation." And finally, since Grelling Paradox can be obtained from Russell's Paradox, in the way we showed above, it follows that both these paradoxes can be used to obtain Gödel's result (QUINE, 1976; SMULLYAN, 1994).

On the other hand, the Gödel's Theorem can be obtained from Grelling's result by moving from "Heterological" (it is not *true* of itself) to the "Gödel Heterological" (is not *provable* of itself). Let us detail.

Let *Prv* (y,x) be the semirecursive relation "y is provable of x", where y is the Gödel number of a formula $\alpha(w)$, with w free. Then $Prv(y,x)$ means: "the result of replacing all free occurrences of w in α with the numeral for x is provable, i.e. $\vdash \alpha(\boldsymbol{x})$". Being semirecursive, it has the form (Ez) $P(y,x,z)$, with $P(y,x,z)$ recursive. Let Θ(y,x,z) be the formula expressing $P(y,x,z)$ in Lpa. Now let us take the relation $\sim$ *Prv* (w,w) and *Form*: $\neg$\$z Θ(w,w,z) be the formula expressing it in Lpa. Let k be the Gödel number of *Form* and G be the sentence $\neg$\$z $\Theta(\boldsymbol{k},\boldsymbol{k},$z). The sentence G is just the undecidable Gödelian sentence for PA, derivable *via* Grelling Paradox. Let us argue that this is the case.

(a) If PA is consistent, then G is true but not provable in PA.

(*Reductio*). Assume hypothesis and that $\vdash$G, i.e. $Prv(k,k)$ holds, and then $P(k,k,n)$holds for some n, and therefore $\vdash\ \Theta(\boldsymbol{k},\boldsymbol{k},\boldsymbol{n})$, hence $\vdash$\$z $\Theta(\boldsymbol{k},\boldsymbol{k},$z) also holds. But since $\vdash$G, by assumption, i.e. $\vdash\neg$\$z $\Theta(\boldsymbol{k},\boldsymbol{k},$z), we derived an inconsistency. So, G is not provable, i.e. *Form* is not provable of itself. But this is what G itself says, and then G is true.

(b) If PA is ω-consistent, then $\neg$G is not provable in PA

Since G is true, $\neg$G is false. Hence $\neg$G is the false Σ_1 -sentence \$z $\Theta(\boldsymbol{k},\boldsymbol{k},$z); it follows, by (*Eq2*) Section 1, that it is not provable in PA, if PA is ω-consistent.

Gödel's Theorem goes into many logical and philosophical arguments, such as the realism-antirealism dispute and the Lucas-Penrose Argument. Actually, in this last case a "computational" form of this theorem is the key point: the Turing's form of Gödel's Theorem. Let us see this fact.

4. KLEENE'S FORM AND TURING'S FORM OF GÖDEL'S THEOREM

Both forms of Gödel's Theorem are based on an application of the formal self-reference, called diagonalization.

Kleene's generalized form of Gödel's Theorem is formulated in terms of the primitive recursive predicate $T(z,x,y)$. Let $P(x)$ be the predicate $(y)\sim T(x,x,y)$ and α(x) be the formula defining it in a formal system S, i.e. for any n:

$P(n)$ holds iff $\alpha(\boldsymbol{n})$ is true

S is correct for α(x) if for any n holds: $\vdash\ \alpha(\boldsymbol{n}) \rightarrow P(n)$, and S is complete for α(x) if the converse is the case: $P(n) \rightarrow\vdash\ \alpha(\boldsymbol{n})$.

Kleene's generalized form. If α(x) expresses $(y)\sim T(x,x,y)$, then a number f can be found such that $(y)\sim T(f,f,y)$ and $\nvdash \alpha(\boldsymbol{f})$.

Recall the proof. In Kleene's terms, for any formula α(x) a primitive recursive predicate $R(x,y)$ can be found such that

(1) $(Ey)R(x,y) \leftrightarrow \ \vdash \alpha(\boldsymbol{x})$

Let f be the index of $(Ey)R(x,y)$ in an enumeration, i.e.

(2) $(Ey)R(x,y) \ \leftrightarrow\ (Ey)T(f,x,y)$

For $x=f$ (diagonalization!) we have:

(3) $(Ey)R(f,y) \ \leftrightarrow\ (Ey)T(f,f,y)$

Suppose S is correct for α(x), i.e.

(4) $\vdash \alpha(\boldsymbol{f}) \ \rightarrow (y)\sim T(f,f,y)$, and then

(5) $(y)\sim T(f,f,y) \leftrightarrow \sim(Ey)T(f,f,y)$ (by logic) $\leftrightarrow\sim(Ey) R(f,y)$ (by (3)) $\leftrightarrow \nvdash \alpha(\boldsymbol{f})$ (by (1) for $x=f$.

Hence if S is correct for α(x), then S is not complete for α(x) (KLEENE, 1952).

Remark. There are proofs of Kleene's forms of Gödel's Theorem not using the idea of diagonalization *within* PA, but the diagonalization occurs in a new form *outside* PA (SMULLYAN, 1993).

The Turing's form of Gödel's Theorem can be obtained *via* Kleene's form, by interpreting the Kleene's predicate $T(z,x,y)$ in the following terms: z is the Gödel number of a Turing machine Cz and y is the Gödel number of a computation at input x. Then the predicate $(Ey)T(z,x,y)$ says that Cz halts at input x, and its negation $(y)\sim T(z,x,y)$ that it does not halt at input x.

To get the Turing's form of Gödel's Theorem let us firstly slightly modify Kleene's form. Suppose now that $\alpha(\boldsymbol{q,n})$ expresses the predicate $P(q,n)$: $(y)\sim T(q,n,y)$ and that the following holds: $P(q,n)$ iff $Cq(n)$ does not halt. The definitions of correctness and completeness of S for $\alpha(\boldsymbol{q,n})$ can be correspondingly formulated.

Kleene's generalized form (a variant). If $\alpha(\boldsymbol{q,n})$ expresses the predicate $P(q,n)$, then there is a number k such that $Ck(k)$ does not halt and $\nvdash \alpha(\boldsymbol{k,k})$.

Now, if we take $\alpha(\boldsymbol{q,n})$ to be the pair (q,n) and the semirecursive set *Th* of the theorems of *S* to be the set of pairs on which a Turing machine halts, then the following form of Gödel's Theorem is obtained (PENROSE, 1994).

Turing's form of Gödel's Theorem. Let A be a Turing machine such that the following holds: If for an input (q,n) A halts, then $Cq(n)$ does not halt. Then there is a number k such that $Ck(k)$ does not halt and A does not halt on (k,k).

The Turing's form of Gödel's Theorem is the key scientific result on which the Argument Lucas/Penrose is focused, in order to argue that the human mathematical thinking transcends the powers of any Turing machine (PENROSE, 1994).

Let us finally consider a very special case of approaching the idea of reflexivity of a language: Wittgenstein.

5. WITTGENSTEIN AND THE SUPPRESSING OF ASSERTED REFLEXIVITY OF LANGUAGE

Wittgenstein's early philosophy of *Tractatus* lies in the Kantian horizon of the transcendental philosophy. Of course, the radical changing of the conceptual frame consists of the passing from the logical form of the object-*consciousness* to the logical form of object-*description*. So, the transcendental difference in *Tractatus* is that between what can be described (stated, experienced) and the conditions of the possibility of experience: the "logical form" of the language, identical to the "logical form" of the world (STEGMÜLLER, 1989). Concisely expressed, this difference is that between "what can be said" (*was sich sagen lässt*) and "what only shows itself" (*was sich nur zeigt*). The logical form of language/world belongs to this last province. Which are the consequences of this expulsion of the conditions of the possibility of experience from the domain of what can be stated? Simply, any attempt to speak about them rejects the respective speaking as being nonsensical, as the aphorism 6.54 of *Tractatus* states:

> My propositions [*Sätze*] serve as elucidations in the following way: anyone who understands me eventually recognizes them as nonsensical when he has used them –as steps– to climb up beyond them. (He must, so to speak, throw away the ladder [*Leiter*] after he has climbed up it.)

He must transcend these propositions, and then he will see the world aright (WITTGENSTEIN, 1963; WITTGENSTEIN, 1988).

Why are these "Leiter Sätze" of transcendental semantic "nonsensical"? Two aphorisms of *Tractatus* try to give a lapidary answer.

Firstly, in 4.2 we find:

> Propositions […] cannot represent […] logical form. In order to be able to represent logical form, we should be able to station ourselves with propositions somewhere outside logic, that is to say outside the world.

This aphorism seems to be a direct consequence of the Wittgenstein's transcendental critique of sense: what happens if we claim to formulate an *empirical* criterion of sense for *transcendental* conditions of the possibility of experience? This is a priori impossible, since we should adopt a stand point on the outside of logical form, i.e. outside of the condition of possibility of expressing itself. Therefore, the logical form only shows itself (CARNAP, 1963).

This is one way on which Wittgenstein suppresses the reflexivity of language in expressing the transcendental character of logical form. But those seven theses of his *Tractatus* just speak about the logical form. And then nonsensicality (*Unsinnigkeit*) of his metaphysics follows immediately.

And secondly, according to 3.332, by Russell's Theory of Types,

> [n]o proposition can make a statement about itself, because a propositional sign cannot be contained in itself […].

Hence the *Sätze* of *Tractatus* are nonsensical since they violate the limit of what can be said, a limit traced by Russell's Theory of Types. Is this, we ask ourselves, a strong argument against this kind of language reflexivity? If we carefully read the construction of Russell's Paradox, we can see that some ingredients are necessary in order to derive the paradox: an idea of self-reference, in the form $(x \in x)$, a universal quantifier, (x), and a concept of *negation*, ~, occurring in the expression $\sim(x \in x)$. What happens if we erase the negation? In this case the paradox cannot be constructed, even the self-reference and the universal quantifier are kept, and then the reflexivity of language can consistently be preserved (DRAGHICI, 2005).

Finally, let us ask ourselves why Wittgenstein's philosophy seems to be so strange, ending up by declaring itself as nonsensical? The story is long. We confine ourselves to consider that essentially this fact is due to the main ideas of his philosophy: the thesis "meaning is use", the transcendental standpoint, his positivism and the rejection of Hilbert's metamathematic (i.e. the rejecting of the distinction object language – metalanguage) (WITTGENSTEIN, 1984a,b,c).

6. CONCLUSIONS

The reflexivity of a language represents its ability to speak about its own expressions. A special sort of this property is represented by the self-referential constructions, the most striking example being that of the Liar Paradox. In the history of philosophy the attitudes towards such constructions were very different. In the preceding considerations I took two paradigmatic cases: Gödel and Wittgenstein. In the first case, the cost of solving the Liar Paradox is the introduction of some limitative results, either in the form of incompleteness theorems (Gödel's results), or in the form of separating a language in object language and metalanguage (Tarski's result). Both these sorts of results are treated as being based on formal self-referentiality called *diagonalization*. On the other hand, the Gödel's result is the basis of an important logical, mathematical and philosophical argument, the Lucas/Penrose Argument, in which, again, the idea of diagonalization is the main means of proof. Actually, this argument is based on a Turing's form of Gödel's Theorem. Section 3 shows how it can be derived, *via* Kleene's generalized form of Gödel's result. In the second case, Russell's Paradox (a diagonal construction too!) represents a key argument for rejecting the propositions of metaphysics. In order to connect the two cases, section 2 shows how the three paradoxes (Liar, Grelling and Russell's) are mutually deducible, and therefore how Gödel's Incompleteness Theorem can be derived from Russell's Paradox. Finally, section 4 shows how the self-referentiality of a language works in Wittgenstein's philosophy of *Tractatus* in order to annihilate, *via* Russell's Paradox, its own horizon of conceptualization. And that this stance is not independent on other major philosophical assumptions.

References

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

BOOLOS, G., BURGESS, JP & JEFFREY, RC (2002) *Computability and Logic*, Fourth Ed. Cambridge University Press, Cambridge.

CARNAP, R., FRANK, P & SCHLICK, M (1934) *Logische Syntax der Sprache*, Springer-Verlag Berlin Heidelberg, Heidelberg.

DRAGHICI, V. (2005) "Leiter-Sätze", *Proceedings of the International Conference Humanities and Social Sciences Today. Classical and Contemporary Issues*. PROUniversitaria Publishing House, Bucuresti.

GÖDEL, K. (1931) Über formal unentscheibare Sätze der Principia Mathematica und verwandter Systeme I, *Monatshefte für Mathematik und Physik*, 38(1), 173-198.

GÖDEL, K. (1986) *Collected Works*, vol I. Oxford University Press, Oxford.

KLEENE, S.C. (1952) *Introduction to Metamathematics*, Van Nostrand, Princeton.

PENROSE, R. (1994) *Shadows of the Mind*. Oxford University Press, Oxford.

QUINE, W.O. (1976) *The Ways of Paradox and other Essays*. Harvard University Press, Cambridge.

SMORYNSKI, C. (1977) The incompleteness theorems, In: J. Barwise (ed). Handbook of Mathematical Logic. North-Holland, New York, p.821-865.

SMULLYAN, R.M. (1993) *Recursion Theory for Metamathematics*. Oxford University Press, Oxford.

SMULLYAN, R.M. (1994) *Diagonalization and Self-Reference*. Clarendon Press, Oxford.

STEGMÜLLER, W. (1989) *Hauptströmmungen der Gegenwartsphilosophie*. Alfred Kröner Verlag, Stuttgart.

TARSKI, A., (1936) Der Wahrheitsbegriff in den formalisierten Sprachen, *Studia Philosophica*, 1, pp. 261-405.

TARSKI, A. (1944) The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, 4(3),p. 341-376.

TARSKI, A. (1956) *Logic, Semantics, Metamathematics*. Oxford University Press, Oxford.

WITTGENSTEIN, L. (1988) *Tractatus Logico-Philosophicus*. Routlege, London.

WITTGENSTEIN, L. (1963) *Tractatus logico-philosophicus. Logisch-philosophische Abhandlung*, Suhrkamp Verlag, Berlin.

WITTGENSTEIN, L. (1984a) *Philosophische Grammatik*, Suhrkamp Verlag, Berlin.

WITTGENSTEIN, L. (1984b) *Bemerkungen über die Grundlagen der Mathematik*, Suhrkamp Verlag, Berlin.

WITTGENSTEIN, L. (1984c) *Philosophische Bemerkungen*, Suhrkamp, Berlin.