STUDIA UBB. PHILOSOPHIA, Vol. 64 (2019), Special Issue, pp. 111-118
(RECOMMENDED CITATION)
DOI:10.24193/subbphil.2019.spiss.08

SELF-REFERENCE AND THE LIMITS OF THOUGHT

LUCIAN CONSTANTIN PETRAŞ*

ABSTRACT. Self-reference and the Limits of Thought. This paper explores the connection between the natural language and a formal language from a particular point of view: *self-referential constructions*. Such constructions lead to some kind of *limits of thought*, either in the form of paradoxical constructions (Liar-type or Grelling-type), or in the form of the so called limitative theorems in mathematical logic (e.g. Gödel's theorem). By deriving Gödel's significant results from paradoxical constructions the limitative character of such self-referential constructions is preserved, but they open the ways for a new representation of a great variety of arguments in the field of logic, mathematics and philosophy.

Keywords: *self-reference, paradox, incompleteness theorems, Gödel, Grelling*

Preliminary

In his celebrated paper [1931][1] K. Gödel showed how to construct a sentence in the language of an appropriate formal system $\mathcal{S}$ such that this sentence is undecidable in $\mathcal{S}$ and, moreover, we can argue that it is true. As Gödel himself says, "[t]he analogy of this argument with the Richard antinomy leaps to the eye. It is closely related to the «Liar» too [...]"[2] and that "[a]ny epistemological antinomy could be used for a similar proof of the existence of undecidable propositions".[3] The following paper explores this Gödel's suggestion, that of deriving his main result from other two paradoxes: Liar Paradox (or Epimenides Paradox) and Grelling Paradox. The whole analysis is based on the idea of self-reference and its aim is to make explicit the idea of some limits of thought.

* PhD candidate, Doctoral School in Philosophy, Faculty of History and Philosophy, Babeş-Bolyai University, Cluj-Napoca, Romania. E-mail: *lucian.petras@ravago.ro*

[1] K. Gödel, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme", *Monatshefte für Mathematik und Physik*, 38, 1931, 173–198; a germ.-engl. ed., cf. K. Gödel, *Collected Works*, vol. I, Oxford University Press, 1986, 144–195 (cited here).

[2] K. Gödel, "Über [...]", 140.

[3] K. Gödel, "Über [...]", footnote 14.

1. An absolute limit generated by paradoxes

Self-reference and semantic paradoxes

In its usual sense, a sentence is called self-referential if it asserts something about itself. Let us give the following examples:

(a) This sentence has only six words.

(b) This sentence is the second example in the list.

(c) This sentence has ten words.

If the pronoun "this" refers to the sentence containing it, then all the sentences in the list are self-referential, (a) and (b) true and (c) a false sentence. Then, as can be seen, by itself a self-referential construction is not necessarily paradoxical. When becomes it a paradox? An answer can be given by a simple inspection of two semantic paradoxes: *Liar Paradox* and *Grelling Paradox*.

Liar Paradox (LP): *This sentence is not true.*

Why is LP paradoxical? Suppose LP is true. Then it is true what LP asserts. Hence it is not true. It follows that LP is not true. But being not true, it is not true what it asserts, hence it is true. Therefore, LP is true iff[4] LP is not true. And then LP is paradoxical. Shortly, a sentence *S* is a (semantic) paradox if it is impossible to assign it a truth value.

Grelling Paradox (GP)

This paradox, also known as "heterological paradox", can be derived in the following way. An adjective is called *autological* if it has the property it denotes or, equivalently, it is true of itself, or it does satisfy the property it expresses, respectively; e.g. "short", "polysyllabic", "English". And an adjective is called *heterological* if it has not the property it denotes/ expresses or, equivalently, it is not true of itself; e.g. "long", "monosyllabic", "Romanian". Now, is the adjective "heterological" (abrev. "Het") heterological? We have:

"Het" is heterological iff "Het" has not the property it denotes (by definition of "Het") iff it is not heterological. Again, as in the preceding case, we easily derive a paradox:

G.P. "Het" is heterological iff "Het" is not heterological.

[4] "iff", an abbreviation for "if and only if".

Therefore, in both cases, LP and GP, we have a limit of thought: the suspension of reason. By impossibility to assign them a truth value, this limit, in semantical frame they were constructed, has an absolute character. And as can be observed, the means for constructing such semantic paradoxes are: a semantical term, like: "true", "true of", "satisfy", a notion of negation and the idea of self-reference.

Are these paradoxical constructions relevant in any way? As we argue, they are relevant both mathematical and philosophical.

2. Paradoxes and the incompleteness theorem

2.1. The derivation of Gödel's theorem via LP

2.1.1. A heuristic argument

As Gödel remarks, his celebrated result can be derived from Liar Paradox. A simple heuristic argument shows us how to proceed. Let us consider the following items:

(a) LP. This sentence is not *true*.

(b) G. This sentence is not *provable*.

(c) $\mathcal{S}$ is a *sound* formal system, i.e. a system for which the following holds: If $\vdash \alpha$, then α is true, where "$\vdash$" means "is provable in $\mathcal{S}$" and α is a sentence in the language of $\mathcal{S}$.

An abstract form of Gödel's result is the following.

Gödel's theorem. *If $\mathcal{S}$ is sound, then G is true but not provable in $\mathcal{S}$.*

Argument. Suppose that *G* is provable. Then *G* is false (by (b)), hence *G* is not provable (by (c)). Therefore, *G is not provable* (by *reductio*). But the non provability of *G* is exactly what *G* itself asserts. Then *G* is *true*.

As we saw, the limit imposed by LP is just the nonrationality of the sentence itself. Then, *G* is simply obtained from LP by replacing the semantic term "true" with the syntactic one, "provable". By this move the paradoxical character of LP becomes a rational construction, a scientific result in mathematical logic: *a limiting result* regarding some formal systems. Therefore, by passing from a LP to Gödel's theorem, the limit of thought imposed by LP is converted in a limit of thought as an incompleteness phenomenon (under assumption of soundness of $\mathcal{S}$).

2.1.2. A formal derivation of Gödel's Theorem

As we saw, the Gödel's sentence *G* is self-referential, i.e. it is a sentence asserting its own unprovability. However, a *formal* derivation of Gödel's result supposes a *formal* construction of the sentence *G*. To do that the following means are necessary:

(a) The *coding* of expressions,[5] a way by which our metamathematical assertions turn into recursive functions and relations. For example, let $Pf(y,x)$ be the following metamathematical assertion: "*y* is (the code of) a proof of the formula (with the code) *x*". Its aritmetical counterpart is the following expression:

$$Pf(y,x): \mathit{Prf}\ (y) \wedge x = (y)_{lh(y)},$$

where *Prf* (y) is the primitive recursive relation "*y* is (the code of) a proof", $lh(y)$ is the primitive recursive function "the number of nonvanishing exponents in the factorization of *y*", and $(y)_i$ is the primitive recursive function "the exponent k_i of the prime factor p_i (for $i = 1,2,...$) in the factorization of *y*". Since " = " is a primitive recursive relation and the recursiveness of relations is closed under conjunction $(\wedge)$, it follows that $Pf(y,x)$ is a primitive recursive relation.

(b) The *representability* of recursive functions and the *formal expressibility* of recursive relations within a formal system $\mathcal{S}$.[6] Let us suppose that $Pf(y,x)$ is expressed in $\mathcal{S}$ by the formula Π(y,x).[7]

(c) The *diagonalization* of an expression. If α(x) is a formula of $\mathcal{S}$, containing x free, and *n* is its Gödel number, then $\alpha(\overline{n})$ is called its diagonalization; where $\overline{n}$ is the numeral (or canonical name) for *n*. Intuitively, $\alpha(\overline{n})$ says that α is satisfied by its own code. The following important result concerning the diagonalization, allows us the formal construction of *G*:

Diagonal Lemma (DL). *For any formula* α(x), *with* x *free, there is a sentence G such that*

$$\mathcal{S} \vdash \mathrm{G} \equiv \alpha(\overline{g}),$$

where $\overline{g}$ *is the numeral for g*.[8]

---

[5] The *code* of an expression is also called its *Gödel number*.

[6] In what follows we consider that S is a formal system extending Peano Arithmetic (PA).

[7] In order to distinguish between *intuitive* and *formal* symbols, we render the intuitive symbols using *italics*.

[8] For the proof of DL, comp. Boolos, Burgess and Jeffrey, *Computability and Logic*, Ch. 17, and G. Boolos, *The Logic of Provability*, Ch. 3.

Now, the formal construction of *G* proceeds as follows: we take Π(y,x) the formula expressing formally in $\mathcal{S}$ the primitive recursive relation $Pf(y,x)$, construct the formula α(x): $\forall$y$\neg\Pi$(y,x), equivalently, $\neg\exists$yΠ(y,x), whose intuitive meaning is that there is no proof of the formula whose Gödel number is *x*", and apply DL, i.e.:

$$\mathcal{S} \vdash \mathrm{G} \equiv \neg\exists \mathrm{y}\Pi(\mathrm{y}, \overline{g}).$$

Here *G* is a sentence *equivalent* to a sentence asserting the nonprovability of *G*![9]

Now, the famous Gödel's first incompleteness theorem is as follows.

Gödel's Theorem.

(1) *If* $\mathcal{S}$ *is consistent, then* $\nvdash G$.

(2) *If* $\mathcal{S}$ *is* ω*-consistent, then* $\nvdash \neg G$.

Proof. (1) *Reductio*. Assume hypothesis and that $\vdash G$. Hence there is a proof of *G* with, say, the Gödel number *k*, i.e. $Pf(k,g)$ is true. By formal expressibility of $Pf(y,x)$ it follows that $\vdash \Pi(\overline{k}, \overline{g})$. On the other hand, from $\vdash G$ and the result of DL, it follows that $\vdash \forall \mathrm{y}\neg\Pi(\mathrm{y}, \overline{g})$, from which by predicate calculus and *modus ponens* it follows that $\vdash \neg\Pi(\overline{k}, \overline{g})$, destroying the assumed consistency of $\mathcal{S}$.

(2) *Reductio*. Assume hypothesis and that $\vdash \neg G$, i.e., by result of DL, $\vdash \exists \mathrm{y}\Pi(\mathrm{y}, \overline{g})$. Since ω-consistency does imply consistency, from $\vdash \neg G$ follows that $\nvdash G$, i.e., for any *n*, $Pf(n,g)$ is false. And then, by formal expressibility, it follows that for any *n*, $\vdash \neg\Pi(\overline{n}, \overline{g})$, i.e. $\vdash \neg\Pi(0, \overline{g})$, $\vdash \neg\Pi(\overline{1}, \overline{g})$, $\vdash \neg\Pi(\overline{2}, \overline{g})$,..., destroying the assumed ω-consistency.

Therefore, if $\mathcal{S}$ is ω-consistent, then the sentence *G* is undecidable in $\mathcal{S}$; and since *G* is equivalent to a sentence asserting the nonprovability of *G*, it follows that the sentence *G* is true. Shortly, "true" and "provable" with reference to $\mathcal{S}$ do not coincide!

2.2. The derivation of Gödel's theorem via GP

2.2.1. A formal reconstruction of GP

As we saw above (sect. 1), "Het" means "is not true of itself" or "it does not satisfy the property it denotes". Let HET(x) be the formula expressing in the

[9] If α(x) is the formula $\neg$Tr(x) (where Tr(x) is the truth predicate), then, as can be seen, we obtain the Formal Liar.

language of $\mathcal{S}$, $\mathcal{L}_{\mathcal{S}}$, the intuitive predicate $Het(x)$ (i.e. *x* is heterological) and SAT(x,x) be the formula expressing in $\mathcal{L}_{\mathcal{S}}$ the intuitive predicate $Sat(x,x)$ (i.e. *x* does not satisfy *x*). Then we have the definition

Def. $\mathrm{HET(x)} = \neg\mathrm{SAT(x,x)}$.

Now, the formal derivation of GP is as follows. Let *k* be the Gödel number of the formula HET(x). Then $\mathrm{HET}(\bar{k})$ is the diagonalization of HET(x), and it means "*k* is heterological", or "HET(x) is heterological". So, we have:
$\mathrm{HET}(\bar{k})$ iff $Sat(k,k)$ iff $\mathrm{SAT}(\bar{k},\bar{k})$ iff $\neg\mathrm{HET}(\bar{k})$ (by **Def**), i.e. $\mathrm{HET}(\bar{k})$ iff $\neg\mathrm{HET}(\bar{k})$, and this is the formal GP.

2.2.2. A formal derivation of Gödel's theorem

As in the preceding derivation of *G* from LP, now we replace the semantic notion $Sat(x,x)$ with a syntactic one: *Prov*(x,x): "*x* is provable of *x*".[10] Let PROV(x,x) be the formula expressing it in $\mathcal{L}_{\mathcal{S}}$.

Def. $\mathrm{GHET(x)} = \neg\mathrm{PROV(x,x)}$,

where GHET(x) means "*x* is Gödel-heterological, that is the formula with Gödel number *x* is *not* provable of itself. Let *k* be the Gödel number of GHET(x) and $\mathrm{GHET}(\bar{k})$ be its diagonalization.

Now, Gödel's result, via GP, runs as follows:

Gödel's Theorem. *If* $\mathcal{S}$ *is consistent, then* $\mathrm{GHET(x)}$ *is Gödel-heterological*.

A heuristic argument (Reductio). Assume hypothesis and suppose that GHET(x) is not Gödel-heterological, that is GHET(x) is provable of itself and this means that

(*) $\mathcal{S} \vdash \mathrm{GHET}(\bar{k})$.

From (*) it follows that:

(a) $\mathcal{S} \vdash \mathrm{PROV}(\bar{k},\bar{k})$ (by Def), and

(b) $\mathcal{S} \vdash \mathrm{PROV}(\bar{k},\bar{k})$, since "GHET(x) is provable of itself" is the *negation* of $\neg\mathrm{PROV}(\bar{k},\bar{k})$.

But (a) and (b) contradict the assumed consistency of $\mathcal{S}$. And since $\nvdash \mathrm{GHET}(\bar{k})$, it follows that $\mathrm{GHET}(\bar{k})$ is true.

[10] Some authors derive Gödel's result via GP using Gödel-Grelling formula GG(x): "*x* is not self-applicable"; comp. Boolos, Burgess and Jeffrey, *Computability and Logic*, 228.

Remark. From this heuristic argument the *standard* form of Gödel's Theorem can be derived, by considering the fact that since PROV(x,x) is a Σ_1-formula, then it has the form ∃zPRV(x,x,z), where PRV(x,x,z) is a Σ_0-formula (and then decidable). Now, if GHET(x) is the formula ¬∃zPRV(x,x,z) and *k* is its Gödel number, then $\mathrm{GHET}(\bar{k})$, i.e. $\neg\exists z\mathrm{PRV}(\bar{k},\bar{k},z)$, is the famous Gödel's undecidable sentence. And then the standard form of Gödel's theorem is the following:

(1) If $\mathcal{S}$ is consistent, then $\nvdash$ G (i.e. $\nvdash \mathrm{GHET}(\bar{k})$), and

(2) If $\mathcal{S}$ is ω-consistent, then $\nvdash \neg$G (i.e. $\nvdash \neg\,\mathrm{GHET}(\bar{k})$).[11]

Conclusions

1. The self-referential constructions represent a remarkable tool for exploring the idea of the limit of thought, either in its absolute form of paradoxes in usual languages (e.g. Liar Paradox and Grelling Paradox), or in its logical and mathematical sense of limitative theorems, with reference to formalized languages. The derivation of Gödel's theorem by reconstructing and reinterpreting some semantic paradoxes makes also explicit the way in which a limitative result generated by paradoxical constructions preserve the limitative character by its rational conversion in an undecidability result.

2. By using the diagonalization lemma we get the means for constructing *formal* self-referential structures, expressing thus a great variety of limits of thought in the form of fixed-point sentences.

3. The existence of such limits of thought has a great impact on contemporary philosophical theorizing, either in the form of the sophisticated Lucas/Penrose Argument and its associated topic of the realism-antirealism controversy, or in the form of destruction of the far-reaching metamathematical and philosophical programs (e.g. Hilbert's, Frege's, Wittgenstein's and Heidegger's).

[11] For proof and details, comp. V. Drăghici, "The reflexivity of a language", *Cultural and Linguistic Communication*, Vol. 8, Issue 3, 2018, 218–223.

BIBLIOGRAPHY

Boolos, G., *The Logic of Provability*, Cambridge UP, 1993.

Boolos, G., Burgess, J. P., Jeffrey, R. C., *Computability and Logic*, Fourth Ed., Cambridge UP, 2002.

Gödel, K., "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme", *Monatshefte für Mathematik und Physik*, 38, 1931, 173–198; a germ.-engl. edition, comp. K. Gödel, *Collected Works*, Vol. I, Oxford University Press, 1986, 144–195.

Kleene, S. C., *Introduction to Metamathematics*, North-Holland Publishing Co., Amsterdam, 1964.

Mendelson, E., *Introduction to Mathematical Logic*, Princeton, 1964.

Drăghici, V., "The reflexivity of a language". *Cultural and Linguistic Communication*, Vol. 8, Issue 3, 2018, 218–223.