Cogito ergo sum non machina!

About Gödel's First Incompleteness Theorem and Turing machines.

Ricardo Pereira Tassinari[1]

Philosophy Department of State University of São Paulo - UNESP - Campus Marília

Itala M. Loffredo D'Ottaviano

Group for Theoretical and Applied Logic, Centre for Logic, Epistemology and the History of Science and Philosophy Department of the State University of Campinas - UNICAMP

Abstract

The aim of this paper is to argue about the impossibility of constructing a complete formal theory or a complete Turing machines' algorithm that represent the human capacity of recognizing mathematical truths. More specifically, based on a direct argument from Gödel's First Incompleteness Theorem, we discuss the impossibility of constructing a complete formal theory or a complete Turing machines' algorithm to the human capacity of recognition of first-order arithmetical truths and so of mathematical truths in general.

1. Introduction

Know Yourself!
Delphos' Oracle

Could we construct models, by using formal theories or algorithms, in order to completely express the capacity of (some) human beings in identifying arithmetical truths? Some authors, like Lucas (1961) and Penrose (1989 e 1995), insist in a negative answer. Penrose (1989 e 1995) argues, from an analysis of the Halt Problem, that the form of the human thinking is not mechanical. Lucas (1961) argues the same thing, but in a non-direct way; he constructs a 'schema of refutation': given any computational program that simulates the human capacity of recognizing arithmetical truths, he shows

[1] This article corresponds to part of the results of the Doctorate Thesis *Incompleteness and self-organization: on the determination of mathematical and logical truths*, of the first author, under the supervision of the second one, that was presented at the Institute of Philosophy and Human Sciences of the State University of Campinas - UNICAMP, Brazil, in December of 2003.

how to use Gödel's First Incompleteness Theorem in order to find one formula that should be recognized as true by the program, but is not.

Could we present another answer to the initial question, more resumed than the one given by Penrose and more direct than Lucas' one? That is the purpose of this paper.

Our perspective is that of someone who tries to construct a theoretical model to describe one part of the human cognition capacity. We analyze, in a epistemological and metamathematical sense, the form of the human capacity of recognizing of some kinds of first-order arithmetical truths (those that are involved in the demonstration of Gödel's First Incompleteness Theorem); and then we ask for the possibility of our capacity being completely expressed in a formal theory or completely modeled by an algorithm.

The first difficult to answer the proposed question is to define what is the human capacity of verification of a first-order arithmetical formula. We preliminary understand that the question belongs to the logical-mathematical scope and we assume the definition of a true formula due to Tarski (1936, [see 1983]) or those which are similar and usual in the textbooks. However, note that we do not have, *prima facie*, a definition of what is *the human capacity* of verification of a first-order arithmetical formula.

On the other hand, for the theoretician that asks for what is the human capacity of verification of first-order arithmetical formulae, the question may be initially analyzed from cases in which we can in principle determine the truth of the sentence. For example, the theoretician knows that we can in principle determine any free variable formulae: it is enough to calculate the successor, the addition or the multiplication of the given terms according to the order of the applications of these operations in the given formula and then to verify the equality of the results of the calculus.[2]

It is important to note that if we take the human capacity of verifying a first-order arithmetical formula *in principle*, we exclude the memory and the time limitations for that verification. As we are trying to express completely such capacity by formal theories or computational algorithms, we can suppose having any space and time as needed, since we admit the same thing for the execution of an ideal program or for the set of ideal deductions in a formal theory. The position here is clear: as we deal with the possibility of modeling either by formal theories or by Turing machines' algorithms the

[2] We are considering that the first-order arithmetic's language has the identity symbol, the zero constant and the function symbols of successor, addition and multiplication.

human capacity of verification of formulae, and either there exists *in principle* no limit in the length of the demonstrations in a theory or there exists *in principle* no limit either of memory or of time in a Turing machines' processing, then we assume that *in principle* we have no limitation in space, memory or time to investigate the truth of one formula.

2. The impossibility of a complete axiomatic first-order theory of human capacity of recognizing first-order arithmetical truths

> *Note that the results mentioned in this postscript do not establish any bounds for the powers of human reason, but rather for the potentialities of pure formalism in mathematics.*
>
> *Gödel (1965, pp. 72-73)*

First of all, we can verify if it is possible to construct a formal first-order theory **T** whose theorems are exactly the formulae that we could recognize *in principle* as true in the structure of natural numbers, that we call as usually the *standard model*. For simplicity, we can consider the following convention.

Notation. We will denote by **ψ(A)**, where **A** is a first-order arithmetical formula, that **A** is true in the standard model and this can be recognized by one logician-mathematician.

Note that ψ denotes a metamathematical unary predicate. More than that, we assume that ψ denotes a partial predicate, i.e., ψ may be not defined for all formulae of the formal system in consideration (it is possible that we cannot know every true first-order arithmetical formula, even *in principle*).

We then consider then the following version of Gödel's First Incompleteness Metatheorem[3].

Gödel's First Incompleteness Metatheorem. Given a consistent axiomatic

[3] From here, we will avoid confusion by using the term 'metatheorem' for a theorem in the metalanguage.

formal theory **T**, whose language is an extension of the first-order arithmetical language and in which the recursive function are represented, then there exists a first-order formula $\mathbf{G_T}$ (the *Gödel Formula*), and we can exhibit it, such that:

(1) $\mathbf{G_T}$ is true in the standard model; and

(2) $\mathbf{G_T}$ is not a theorem of **T**.

Note that the Gödel Formula $\mathbf{G_T}$ has the form[4]:

$$\Pi \boldsymbol{v}[\sim \boldsymbol{B}(\boldsymbol{v}, \boldsymbol{S}(\boldsymbol{w}, \boldsymbol{w}))] \ ,$$

where Π is the universal quantifier, $\boldsymbol{v}$ and $\boldsymbol{w}$ are individual variables, $\boldsymbol{B}$ is a binary predicate symbol and $\boldsymbol{S}$ is a binary function symbol, these last two designating respectively the primitive recursive relation $\mathcal{B}$ and the primitive recursive function γ defined by Gödel (1965, respectively in p.57 and p.59), that will be discussed later on. So $\mathbf{G_T}$ is a first-order formula. Note that, as they are defined, $\mathcal{B}$ is a relation between natural numbers and γ is the function from the set of pairs of natural numbers in the set of natural numbers (and not a relation between formulae and a function from the set of pairs of formulae in the set of formulae, as some people usually erroneously think), defined explicitly by composition and primitive recursion from the constant, projection and successor functions.

In his demonstration, Gödel shows: (1) that we can associate a unique natural number to each formula, today known as the *Gödel number of the formula*; (2) that we can associate a unique natural number to each sequence of formulae, today known as the *Gödel number of the sequence* of formulae; (3) that the primitive recursive relation $\mathcal{B}$, that is designated in the formal system by $\boldsymbol{B}$, is such that, given two numbers $\boldsymbol{x}$ and $\boldsymbol{y}$, $\mathcal{B}(\boldsymbol{x}, \boldsymbol{y})$ happens if and only if $\boldsymbol{x}$ is the Gödel number of the sequence that constitutes a demonstration of the formula whose the Gödel number is $\boldsymbol{y}$; and (4) that the primitive recursive function γ, that is designated in the system by $\boldsymbol{S}$, is such that, given two numbers $\boldsymbol{x}$ and $\boldsymbol{y}$, the result $\gamma(\boldsymbol{x}, \boldsymbol{y})$ is the Gödel number of the formula that results of the substitution, in the formula that the Gödel number is $\boldsymbol{x}$, of all free occurrences of the variable $\boldsymbol{w}$ by the term that is the numeral that represents the number $\boldsymbol{y}$.

[4] We use here the same notation as Gödel (1965).

From that and the fact that z_p denotes the numeral that represents the number p in the formal system, Gödel (1965, p.60) concludes:

> Let $U(w)$ be the formula $\Pi v[\sim B(v, S(w, w))]$ and let p be the number of $U(w)$. Now $U(z_p)$ is the formula which results when we replace all free occurrences of w by z_p in the formula whose number is p, and hence has the number $\gamma(p, p)$. Hence if $U(z_p)$ is provable, there is a k such that $k\mathcal{B}\gamma(p, p)$. But since $S(u, v)$ represents $\gamma(p, p)$ and $B(u, v)$ represents $x\mathcal{B}y$, it follows that $B(z_k, S(z_p, z_p))$ is provable. Also, it is a property of our system that if $\Pi\, v\, F(v)$ is provable, then $F(z_l)$ is provable for all l; consequently, if $U(z_p)$ is provable, $\sim B(z_k, S(z_p, z_p))$, as well as $B(z_k, S(z_p, z_p))$, is provable, and the system contains a contradiction. Thus we conclude that $U(z_p)$ cannot be proved unless the system contains a contradiction.

If we interpret the Gödel Formula $\Pi v[\sim B(v, S(w, w))]$, we have that $\Pi v[\sim B(v, S(w, w))]$ occurs if and only if there is no Gödel's number k such that $k\mathcal{B}\gamma(p, p)$, what is equivalent to assert that there is no demonstration, in the considered formal system, of the formula whose the Gödel number is p. But this is the proper Gödel Formula $\Pi v[\sim B(v, S(w, w))]$. So, if the system is consistent, the truth of the Gödel Formula is equivalent to its indemonstrability in the system. Therefore, if the system is consistent, its Gödel's Formula is true and unprovable.

From our comprehension of the demonstration of Gödel's First Incompleteness Metatheorem, we can say that *in principle* if we can recognize that the theory is consistent, then we can recognize that its Gödel's Formula is true.

Note that, as the question here is to determine the complete theory **T** of the human capacity of verification of first-order arithmetical formulae, we can assume that if this theory **T** exists, we could recognize that the axioms of **T** work and then we could recognize that the theory **T** is consistent. So we can say that the human capacity of recognizing first-order arithmetical truths follows the following principle.

Gödel's Self-Overcoming Principle. Given an axiomatic first-order theory **T** about natural numbers, in which we can represent the recursive functions and such that $\psi(\mathbf{A})$ for all axiom **A** of **T**, then there exists a Gödel Formula $\mathbf{G_T}$, and we can *in principle* exhibit it, such that:

(1) $\psi(\mathbf{G_T})$;

(2) $\mathbf{G_T}$ is not a theorem of **T**.

Based on that principle, it follows the answer to our initial question.

First Consequence of Gödel's Self-Overcoming Principle. There exists no axiomatic first-order theory **T** about natural numbers in which we can represent the recursive functions and such that **A** is a theorem of **T** if and only if **ψ(A)**. In other words, there exists no complete axiomatic first-order theory **T** of the formulae that are recognized as true by the human capacity of verification of a first-order arithmetical formula.

In fact, if there exists a theory **T** in the conditions above, then by Gödel's Self-Overcoming Principle there exists a formula $\mathbf{G_T}$ such that $\boldsymbol{\psi}(\mathbf{G_T})$, and is not a theorem of **T**, what contradicts our assertion that **T** is complete, i.e., that **A** is a theorem of **T** if and only if **ψ(A)**.

3. The impossibility of a complete Turing machines' algorithm of human capacity of recognition of first-order arithmetical truths

> *We now define the notion, already discussed, of an* effectively calculable *function of positive integers by identifying it with the notion of a recursive function of positive integers (or of a λ-definable function of positive integers). This definition is thought to be justified by the considerations which follow, so far as positive justification can ever be obtained for the selection of a formal definition to correspond to an intuitive notion.*
>
> *Church (1965, p.100)*

We can now investigate the implications of the First Consequence of Gödel's Self-Overcoming Principle, concerning to the possibility of simulating in a Turing machine the human recognition of first-order arithmetical true formulae.

First, we will consider that there exists a Turing machine **M** that calculates the results of applicating a predicate **P** if and only if **P** is recursive, as it was demonstrated by Turing (1965, p.149) and Church (1965, p.99, [Meta]Theorems XVI-XVII); and we will consider analogously that there exists a Turing machine **M** that calculates the results of application of a partial predicate **P** if, and only if, **P** is partial recursive.

Then, we can consider the following definition and results obtained by Kleene (1965, p.271):

> *Let $P(x_1, ..., x_n)$ be a predicate which may not be defined for all n-tuples of natural numbers as arguments. By a* completion *of P we understand a predicate Q such that, if $P(x_1, ..., x_n)$ is defined, then $Q(x_1, ..., x_n)$ is defined and has the same value, and if $P(x_1, ..., x_n)$ is undefined, then $Q(x_1, ..., x_n)$ is defined. In particular, the completion $P^+(x_1, ..., x_n)$ which is false when $P(x_1, ..., x_n)$ is undefined, and the completion $P^-(x_1, ..., x_n)$ which is true when $P(x_1, ..., x_n)$ is undefined, we call the* positive completion *and* negative completion *of $P(x_1, ..., x_n)$ respectively. (In P and P^+, the "positive parts" coincide; in P and P^-, the "negative parts" coincide.)*
>
> *[META]THEOREM VI. The positive completion $P^+(x_1, ..., x_n)$ of a partial recursive predicate $P(x_1, ..., x_n)$ is expressible in the form $(Ey)R(x_1, ..., x_n, y)$ where R is primitive recursive; and conversely, any predicate expressible in the form $(Ey)R(x_1, ..., x_n, y)$ where R is general recursive is the positive completion $P^+(x_1, ..., x_n)$ of a partial recursive predicate $P(x_1, ..., x_n)$.*

Assertion. If ψ is partial recursive, then there exists an axiomatic first-order theory **T** about natural numbers such that: **ψ(A)** if, and only if, **A** is a theorem of **T**.

In fact, if we denote the Gödel number of the formula **A** by **[A]**, by the above theorem we have that there exists a general partial recursive predicate ***R*** such that ψ^+**([A])**, if and only if, ***EyR*([A], *y*)**; and therefore **ψ(A)** is true if, and only if, ***EyR*([A], *y*)**. Let **T** be the theory whose axioms are the first-order formulae with the form $\mathbf{A}\wedge(x_i= x_i)$ such that ***R*(*i*,[A])**. First, **T** is a first-order theory, since **T** has only first-order formulae. Second, **T** is axiomatic, since there is a recursive procedure that determines the axioms of **T**. So, consider that if we have **ψ(A)**, then there exists *i* such that $\mathbf{A}\wedge(x_i= x_i)$ is an axiom of **T** and so, by the Simplification Rule, we have that **A** is a theorem of **T**. So, if **ψ(A)**, then **A** is a theorem of **T**. Otherwise, if **A** is a theorem of **T**, then **A** can be obtained by inference rules from the axioms of **T**, i.e., from formulae $\mathbf{A_i}$ such that $\psi(\mathbf{A_i})$. If we assume that the human capacity of verification of first-order arithmetical formulae is such that if **A** is one formula that follows by inference rules from formulae $\mathbf{A_i}$ that we *can* recognize as true, i.e. $\psi(\mathbf{A_i})$, then the formula **A** itself *can* be recognized as true, i.e. **ψ(A)**, then we have that if **A** is theorem of **T** then **ψ(A)**. We can conclude that, if ψ is partial recursive, then there exists an axiomatic first-order theory **T** about natural numbers such that: **ψ(A)**, if and only if, **A** is a theorem of **T**. Q.E.D.

From the above Assertion and the First Consequence of Gödel’s Self-

Overcoming Principle, we immediately have the following.

Second Consequence of Gödel's Self-Overcoming Principle. ψ is not partial recursive and therefore the human capacity of recognition of first-order arithmetical truths cannot be completely simulated by a Turing machine.

Hence, certainly, Turing machines cannot satisfy Gödel's Self-Overcoming Principle. This principle if we attribute it to the human being — and it seems that we can do it, since it was discovered by the human being by the analysis of its own thought —makes us consider that minds are not only Turing machines.

4. The impossibility of an axiomatic conservative extension of a complete axiomatic first-order arithmetical theory of the human capacity of recognition of first-order arithmetical truths

In this section, we extend the results obtained in Sections 2 to theories that are axiomatic extensions of first-order theories.

Third Consequence of Gödel's Self-Overcoming Principle. There exists no axiomatic formal theory **T** whose language is an extension of the first-order arithmetical language and such that, for all first-order formula **A**, **A** is a theorem of **T** if and only if **ψ(A)**, i.e., we can recognize that **A** is a true arithmetical formula.

In fact, if there exists a theory **T** in the conditions above, then there exists a recursive primitive predicate $\mathcal{B}$, such that $\mathcal{B}(x, y)$ if and only if x is the Gödel number of the sequence that is a proof in **T** of the formula whose Gödel number is y. If we define the unary predicate **Q** such that **Q(A)** if and only if $Ey\mathcal{B}([A], y)$, then by [Meta]-Theorem VI of Kleene (1965, p. 271), **Q** is recursive partial. Let **V** be the conjunction of **Q** and **P**, where **P** is the recursive predicate such that **P(A)** if and only if **A** is a first-order arithmetical formula. So, **V** is recursive partial and there exists a general recursive partial predicate R such that $V^{+}([A])$ if and only if $EyR([A], y)$. Let T_1 be the theory whose axioms are the formulae of type $A\wedge(x_i=x_i)$, such that $R(i,[A])$. In this case, as we have seen, T_1 is an axiomatic first-order theory, and **A** is a theorem of T_1 if and only if

ψ(A). So, if there is a theory **T** in the conditions enunciated, then there exists an axiomatic first-order theory $\mathbf{T_1}$ such that **A** is a theorem of **T** if and only if **ψ(A).** This contradicts the First Consequence of Gödel's Self-Overcoming Principle and then we have that there exists no axiomatic formal theory **T** whose language is an extension of the first-order arithmetical language and such that, for all first-order formula **A**, **A** is theorem of **T** if and only if **ψ(A)**. Q.E.D.

4. Conclusion

Finally, as the set of first-order mathematical truth that the human being can recognize is a part of the set of all mathematical truths that the human being can recognize, we can conclude that there exists no complete formal theory, or complete Turing machines' algorithm, for the human capacity of recognition the mathematical truths.

5. Bibliographical references

CHURCH, A., *An Unsolvable Problem of Elementary Number Theory*. **In** DAVIS, 1965, pp. 88-107. [Presented to the American Mathematical Society, April 19,1935 and printed at the first time in the **American Journal of Mathematics**, vol. 58, pp.345-363, 1936.]

DAVIS, M., **The Undecidable. Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions.** New York: Raven Press, 1965.

GÖDEL, K., *On Undecidable Proposition of Formal Mathematical Systems*. **In** DAVIS, 1965, pp. 39-74. [Notes on lectures given by Gödel at the Institute for Advanced Study during the spring of 1934 that cover ground quite similar to that covered in Gödel's original 1391 paper.]

KLEENE, S.C., *Recursive Predicates and Quantifiers*. **In** DAVIS, 1965, pp. 254-287. [Reprinted from the **Transactions**, v.53, n.1, pp. 41-73. American Mathematical Society, 1943.]

LUCAS, J.R., **Minds, Machines and Gödel.** In the author homepage http://users.oc.ax.uk/~jrlucas/index.html. [Printed firstly in **Philosophy**, XXXVI, pp.112-127, 1961.]

PENROSE, R., **The Emperor's New Mind: Concerning Computers, Minds and Laws**

of Physics. Oxford: Oxford University Press, 1989.

—, **Shadows of the Mind: a Search for the Missing Science of Consciousness.** Oxford: Oxford University Press, 1995.

TARSKI, A., **Logic, Semantic, Metamathematics.** 2.ed. Indianapolis: Hackett Publishing Co., 1983.

TASSINARI, R.P., **Incompleteness and self-organization: on the determination of mathematical and logical truths.** Doctorate Thesis, supervised by Itala M. Loffredo D'Ottaviano. Campinas: Institute of Philosophy and Human Sciences, State University of Campinas, 2003.

TURING, A.M. *On Computable Numbers, with Application to the Entscheidungsproblem*. **In** DAVIS, 1965, pp. 115-154. [Reprinted from **Proceedings of the London Mathematical Society**, Ser. 2, v. 42, 1936-7, pp. 230-265.]