Gödel, Turing, the Undecidability Results and the Nature of Human Mind

Riccardo Bruni
Università degli Studi di Firenze, Dipartimento di Filosofia
via Bolognese 52, Firenze 50139, Italy

March 28, 2006

Abstract

In this paper Turing and Gödel's standpoints toward the implications of the undecidability results are addressed. In the case of Turing, we show how his account on the issue was deeply connected to his project of actually building a computing machine showing an intelligent behaviour. Furthermore, we argue that his claim that the argument based on the halting problem offers no objection to that very project, is strenghtened by a general view on mathematical resoning and intelligence which has an anti-mechanistic flavour. As to Gödel's position, we reformulate, by enhancing its modal character, an argument that is contained in an unublished paper belonging to the early 1950's which ends in an open conclusion on minds and machines. We finally present Gödel's interpretation of the solution of the famous $P = NP$ problem contained in a letter to von Neumann as a further contribuition in this direction of working out the philosophical significance of certain mathematical achievements.

1 Introduction

Many years have passed since the first attempts to consider the very basic undecidability theorems in computability theory and metamathematics as directly estabilishing a bridge between mathematics and philosophy have appeared in literature. Roughly speaking, these results have been regarded as providing the means for a proof of the non-mechanical nature of human mind. The technical content of the argument and its relation to the mathematical theory are not very deep, and this makes even more striking that it may still be appealing despite all efforts that have been spent to show where the devised inferences break down.

The latest formulation of such an argument is due to Roger Penrose in his [12, 13]. As to this particular version, we have now come to what seems like a disproof of it, due to different although relatable analyses made by Pavel Pudlák [14], Stuart Shapiro [15] and Per Lindström [8, 9]. It turned out that Penrose's argument is either essentially relying on ambiguous notions (that is, the concept of 'unassailable mathematical truth' and that of 'a formal system which encapsulates all mathematical means of proofs accessible to human mind'), or it is wrong under a plausible definition of them. The failure of Penrose's gödelian argument seriously affects his project of a scientific investigation on the true nature of human consciousness as based on mathematical, physical and philosophical considerations.

Since in our opinion these analyses have so far settled the matter, time seems right to us for a survey covering a more historical aspect of the issue which has somewhat been left unnoticed.

An inspection of the available sources reveals in fact that both Turing and Gödel had already considered similar implications of the undecidability results. The motivations that moved Turing and Gödel to direct their attention to them were different and in some sense opposed on to the other. Turing aimed at investigating intelligent behaviour as applied to the action of mechanical devices, while Gödel was in search for a rigorous argument that could confirm his beliefs concerning the status of mathematical concepts and the nature of mathematical reasoning. In both these cases, their analyses turned out to be surprisingly meaningful to Turing and Gödel's 'programs' in the fields of early Artificial Intelligence projects and in the philosophy of mathematics respectively.

Reconsidering the whole issue from their perspective may in turn help to do justice to it, and it may even give useful suggestions for some second thought concerning its implications.[1]

2 Turing and the "Mathematical Objection"

In a famous paper published on *Mind* in 1950, Turing discussed an argument based on the undecidability results as providing a "mathematical objection" to the idea of a machine that may successfully partecipate to what he called the Imitation Game, which is currently known as Turing's Test. It was not the first time that Turing was trying to tackle such an objection, but in this case he gave a particularly clear formulation of it.

The argument[2] applies to any (universal) machine $\mathsf{M}(x, y)$ which is supposed to answer questions concerning the behaviour of all (unary) machines (in a given enumeration) in such a way that $\mathsf{M}(m, n) = \texttt{YES}$ iff $\mathsf{M}_m(n) = \texttt{YES}$, and $\mathsf{M}(m, n) = \texttt{NO}$ otherwise. By bringing to contradiction the existence of such an M in the expected way,[3] one would be led to the conclusion that either such an

[1] It seems unnecessary instead to spend some words about the up-to-date character of the problem in question. Despite all disproofs in fact, new attempts to restate the argument, even though with a critical attitude toward Penrose's, quickly arose (see [1]). Even in this case, the (brief) reconstruction of Gödel and Turing's views in particular seems questionable however, if not misleading.

[2] See [17, pp. 444–445].

[3] As usual, it would suffices to define a diagonal machine $\mathsf{D}_\mathsf{M}(x)$ such that $\mathsf{D}_\mathsf{M}(n) = \texttt{NO}$ in case $\mathsf{M}(n, n) = \texttt{YES}$ and $\mathsf{D}_\mathsf{M}(n) = \texttt{YES}$ otherwise, and then try to compute the values of $\mathsf{M}(d, d)$ for a description d of D_M.

M is unable to give always the correct answer to all questions, or that it may sometimes fail to give any answer. But this in turn would make it quite easy for an interrogator to distinguish between a machine and a human player, which would not be subject to the same constraints and could go freely in search for an appropriate method and hopefully solve the problem in question. Thus, if Turing's Test is recognized as a suitable way to measure intelligence, machines cannot show an intelligent behaviour.

Turing's way out had been developed since 1947, when the british mathematician had directed his attention to this objection for the first time.[4] The first half of Turing's answer is not refined philosophically speaking, since it rests, as it may be expected, on the crucial requirement for the whole argument to apply, namely that the logical systems or the machine involved be consistent. The conclusion is then rejected by simply stressing that consistency is not an essential feature of intelligence.

As to its second half, Turing's analysis becomes more interesting. In fact, he points out that this kind of argument always applies to machines which have not been suitably trained, where, on the contrary, the abilities of a human mathematician can be viewed as the result of a continuous and stimulating interaction with other members of the community to which he belongs, including proper education.

So, while dropping the requirement of consistency may be sufficient by itself to escape the argument, Turing thought it necessary to do more in order to pursue the ambitious purpose of actually developing intelligent machineries, as he strenghtfully aimed at since the second half of the 1940's. The key idea was the possibility for a machine to act on its own istructions. This action had to be refined according to a specific training which needed to be directed from the outside on the basis of the application of suitably modified educational methods. As part of this project, Turing described, for example, an experiment in training an "unorganized" machine, that is a machinery built with no specific purpose, by means of external stimula of rewards and punishments.[5]

Since Turing never revealed a particular inclination to purely philosophical issues, it may come as a surprise that these pioneristic researches could be viewed as based on an analysis of mathematical reasoning and intelligence in general. It's even more surprising, since it contrasts with what is usually acknowledged to Turing,[6] that this analysis entails a non-mechanistic approach to human mind.

Turing's analysis of mathematical reasoning is contained in a section of his work on ordinal logics [16], and it is thought as providing a conceptual framework within which the results obtained by his logical investigation could be usefully discussed.

Mathematical reasoning is the result of two different faculties, namely *intuition* (which allows us to produce judgements on the truth/falsity of mathemat-

[4] See [18, pp. 87–88].

[5] See [19, pp. 122-125]. Even though he considered the result unsatisfactory since the whole process was not sufficiently close to the one which applies in the case of a child, Turing reported to have succeded in "organizing" such a device into a universal machine.

[6] See, e.g., [7, 1].

ical statements without any conscious train of thoughts), and *ingenuity* (which is responsible for the construction of proofs for intuitive judgements). With the development of formal logic it became clear that no one of these two faculties could be dispensed with. In particular, it follows from Gödel's theorems that it is not possible to reduce intuition to ingenuity, that is to reduce mathematical activity to the choice within a given formal system of axioms of those steps which are the most efficient ones to build the proof of a given statement. On the contrary, it can be expected for a logical investigation only to give a more precise shape to the intuitive (that is, the non-mechanical) steps, without possibly eliminate them entirely.[7] It follows that mathematical knowledge is the result of a combination of mechanical and non-mechanical forms of reasoning.

In his later papers, where the case of intelligence in general is concerned, a similar scheme seems to apply.

In this case Turing speaks[8] of a need of both *discipline*, which is the ability of obeying orders and which is at best exemplified by the behaviour of an ordinary machine, and *initiative*. Even with respect to these two features, a purely mechanical approach would not allow an exaustive analysis. If the investigation is confined to the most general case, that is to those strategies which are applied to seek for a solution of existential problems,[9] it is true that it could be possible to account for a certain amount of cases by starting from a suitable logical system, and by relying our search on the results concerning the problem known as the extraction of programs from formal proofs.

In order to deal with *all* problems of this sort, however, it is necessary to think to more complicated processes, which are not trivially reducible to purely formal methods.[10] This makes it necessary a more radical turn: to change the kind of machines to be used for this task, and to build new ones on the basis of an appropriate analysis of those processes which are responsible for the turning of the child mind into the adult one.

3 Gödel's Modal Argument

While Turing addressed the "mathematical objection" as it actually has been used since then[11] (namely, as an argument showing that the human mind has

[7] In the case of Turing's investigation on ordinal logics for example, given a complete logic Λ (that is, in the terminology of [16], a function such that the collection $(\Lambda)o$, for any ordinal notation o, contains all true Π^0_2 statements), intuition is needed only to verify that a given expression is a notation for a constructive ordinal.

[8] See [19, pp. 125].

[9] Turing'opinion concerning statements of existential form as providing the most comprehensive collection of problems, was based (see [19, p. 126]) on a claim concerning the possibility of reducing all other forms of problem to this one via arithmetization.

[10] In particular, Turing indicated two additional searches for a solution, namely a *genetical search*, which consists in a process of genetical recombination so to obtain a new one which may result to be more suitable for the solution of the given problem, and an *intellectual search*, which is the one that results from the combined action of all the members of the community.

[11] In his retrospective article [11], the English philosopher J. R. Lucas, who is often wrongly credited as the first who provided the objection in question, stated that his original writing [10]

not a mechanical nature and that there can be no machine equivalent to it), Gödel arrived to it from a completely independent direction. In particular, he thought his version of the inference to help emphasizing the phenomenon of the "inexaustibility of mathematics", as he referred to it.

As he said in a lecture he delivered in 1951, the theorem on the undecidability of the sentence expressing consistency in formal mathematical systems "*makes it impossible that someone should set up a well-defined system of axioms and rules and consistently make the following assertion about it: All of these axioms and rules I perceive (with mathematical certitude) to be correct, and moreover I believe that they contain all of mathematics*".[12]

Recently, Bringsjord and Arkoudas in their [1] have presented a modal argument to show that computationalism (according to which, the human mind 'is' a Turing machine) is false. Further, even in the case of Penrose's latest formulation of the inference it is natural to give a modal reading of the informal notions used therein.[13] Then, in order to make a comparison more fruitful, it might be useful to present also Gödel's analysis of the basic inference by enhancing its modal flavour. Provided that, Gödel's argument goes as follows.

Assume $\mathsf{K}_M\alpha$ to mean that α is known with mathematical certitude. Then, it seems natural to assume that

$$\vdash \alpha \Rightarrow \mathsf{K}_M\alpha$$

holds (where the provability symbol should be understood in a broad and informal sense).

It follows that, for any given formal system $\mathcal{F}$, $\mathsf{K}_M(Corr_{\mathcal{F}} \rightarrow Con_{\mathcal{F}})$ where $Corr_{\mathcal{F}}$ and $Con_{\mathcal{F}}$ are the statements representing correctness and consistency of $\mathcal{F}$ respectively.[14] Assume K_M to be closed under *modus ponens*.[15] By the undecidability of $Con_{\mathcal{F}}$, we can conclude that if $\mathsf{K}_M Corr_{\mathcal{F}}$ then there exists a certain formula β such that $(\mathsf{K}_M\beta \wedge \neg\mathsf{Teor}_{\mathcal{F}}\beta)$ holds (where $\mathsf{Teor}_{\mathcal{F}}\alpha$ has an obvious meaning). This finally yields $\mathsf{K}_M\neg(\mathsf{K}_M\alpha \rightarrow \mathsf{Teor}_{\mathcal{F}}\alpha)$ for a generic α, from which it follows the desired conclusion $\neg\mathsf{K}_M(\mathsf{K}_M\alpha \rightarrow \mathsf{Teor}_{\mathcal{F}}\alpha)$ provided we have assumed $\neg(\mathsf{K}_M\alpha \wedge \mathsf{K}_M\neg\alpha)$.[16]

Suppose now that for a certain formal system $\mathcal{F}^*$, we have

$$\mathsf{K}_M\alpha := \mathsf{Teor}_{\mathcal{F}^*}\alpha \quad (*)$$

had been conceived as a reply to Turing's position on machine and intelligence as formulated in [17]. It seems thus natural to infer that Lucas didn't realize that the paper he argued against, already contained a formulation of the very same argument (even clearer than Lucas' own), and Turing's response to it.

[12] See [5, p. 309].

[13] See in particular Lindström [8, 9] and Shapiro [15] on this.

[14] As it is customary, $Corr_{\mathcal{F}}$ must be understood as some form or another of a reflection principle. Then, the implication of the system consistency would be provable (even in a formal framework) and then knowable according to the assumption above.

[15] In symbols, $\mathsf{K}_M(\alpha \rightarrow \beta) \rightarrow (\mathsf{K}_M\alpha \rightarrow \mathsf{K}_M\beta)$ holds.

[16] Our conclusion seems different than Gödel's since, under the assumption that we know with mathematical certainity that a system $\mathcal{F}$ is correct, it shows that we cannot know with the same certainity that this system contains all of mathematics. Gödel's conclusion instead dealt with two modalities, knowledge (with mathematical certitude) and belief, the latter of which we've ommited for the sake of simplicity. We would have obtained a literal translation

which provides a definition for K_M as based on two assumptions, namely (i) that to know a mathematical proposition with certitude means to have a proof of it, and (ii) that there exists a certain formal system which comprises all the commonly acceptable means of proof that are accessible to human reasoning.

Then $\neg\mathsf{K}_M Corr_{\mathcal{F}^*}$ must hold. In Gödel's own words, it would follow that "the human mind (in the realm of pure mathematics) *is* equivalent to a finite machine that, however, is unable to understand completely its own functioning".[17] This conclusion has an immediate philosophical significance for the view entailed by (i) and (ii) above which would turn out to be, so to say, not self-contained since, under these assumptions, there would be no mathematically grounded justification for our belief in the validity of the commonly accepted inferences of deductive reasoning.

Furthermore, consider the collection $\mathbf{M}$ of all mathematical propositions which hold in an absolute sense. Let instead $\mathbf{K}$ be the class of all statements α such that $\mathsf{K}_M\alpha$ holds, and let us assume moreover $\mathbf{K} \subseteq \mathbf{M}$. Then, if $(*)$ holds for a given formal system, it follows that this inclusion is proper, namely there are mathematical truths which are 'inaccessible' to all the mathematical means of proof that the human mind can conceive. If on the contrary $(*)$ fails for all $\mathcal{F}$, it would follow for both $\mathbf{K}$ and $\mathbf{M}$ that these collections cannot be recursively enumerated where it remains possible for the above inclusion to be proper. This leads to Gödel's disjunctive conclusion[18] that either mathematics cannot be comprised in any finite rule, or there exist absolutely unsolvable mathematical problems, or both these alternatives are the case.[19]

Gödel's own solution of this disjunction is known. On the one hand, he thought that "Hilbert was right" in rejecting the existence of absolutely unsolvable mathematical problems among number-theoretic ones.[20] On the other, Gödel was a supporter of an anti-mechanistic view of human reason. In a late and obscure remark,[21] he presented his position as based on the idea that (i) the human mind may be capable of thinking to infinitely many things and possibly even non-denumerably many, and that (ii) the way in which the abstract concepts enter human understanding seems of a *procedural* character (more and more abstract concept become accessible in the course of the mind development), but of a non-mechanical nature (a systematic method to actualize this development would result in a non-recursive arithmetic function).

---

of Gödel's ending by introducing another modal operator B, and equivalently assuming that

$$\neg(\mathsf{K}_M\alpha \wedge \mathsf{B}\neg\alpha)$$

holds (which simply says that belief cannot contrast mathematical knowledge, since the latter has a stronger epistemic force).

[17] [5, pp. 309–310].

[18] See [5, p. 310].

[19] Due to an unpublished refinement of Gödel's theorem on arithmetical equivalents of the formally undecidable sentences (see [4]), the unsolvable problems would moreover have the form of diophantine statements of the type $\forall\vec{x}\exists\vec{y}P(\vec{x},\vec{y}) = 0$, where P is a polynomial with integer coefficients.

[20] See [20, p. 325]. Gödel's reference to Hilbert was motivated by the rationalistic attitude of the latter, as condensed in his famous slogan "In mathematics there's no *ignorabimus*".

[21] See [3, p. 306] and [20, pp. 325–326].

Interestingly, in his Gibbs lecture Gödel argued that the conclusion of his argument could turn out to support some form or another of a Platonistic approach to mathematics: the existence of absolutely unsolvable problems to the human mind would go against the idea that mathematical concepts are our own construction since a creator knows all the properties of what he has created. Thus, it would be required to admit that these concepts, or at least some part of them, have an existence which is independent from our definitions and constructions.[22]

4 From Computability to Complexity

Obviously one may object that nothing like the disclosing of the true nature of human reasoning is *really* the purpose of the argument we've just surveyed. The crucial thesis one is dealing with is in fact comprised by statement $(*)$ above, which should be read as expressing the existence of a machine replicating the activity of human reason in the domain of mathematics when the latter is regarded extensionally. If compared to more recent attempts, the importance of Gödel's argument is primarily due to the refined analysis of the, so to say, linguistic side of the problem. Namely, it shows that there're reasonable interpretations for the concepts involved that lead to a meaningful and rigorously drawn conclusion which connects the mathematical investigation of the foundations to certain conceptual issues. But significantly, the conclusion provide no solution in the latter direction.

Indeed, the typical feature of this sort of arguments is their abstract character: they are concerned with an idealized human mathematician, whose actions are compared with a mathematical model of mechanical computation. This brings to an unsatisfactory ending which can be expressed philosophically by saying that the whole discussion is indifferent to how refined the assumed standpoint concerning the human mind might be. In this sense, the conclusion that can be drawn by Turing's analysis of the inference, which we would summarize by saying that the goals of a scientific research in the field of Artificial Intelligence turns out to be independent from the solution to the problem concerning the nature of human reasoning, is not surprising.

Then, one may try to refine the investingation by introducing some very basic restriction.

An attempt could be the one which starts from a view of human reasoning as the sum of actual processes which, like all kind of processes, are subject to certain practical constraints. One would consequently assume that only those processes which are feasible matter. Then, it is natural on the mathematical side to shift from computability to complexity issues.

Gödel is known to be among the firsts who adopted a related perspective. He did that in a letter to von Neumann dated 20 March 1956, which became

[22] This argument is not that conclusive. One may argue against it, for example, by emphasizing that, due to Gödel's own theorems, the identification of mathematical concepts with their symbolic representation in a formal language does not lead to omniscience.

known for containing one of the first formulation of the $P = NP$ problem.[23]

Indeed, Gödel questioned his correspondent about the following problem: to consider a Turing machine M that allows to decide, for every $n \in \mathbf{N}$ and every formula F of first order predicate logic, whether F has a proof of lenght n (n being the number of its symbols); then, if by $\psi(n, F)$ we indicate the number of steps required by the machine to reach a decision, Gödel's question was whether it is possible for

$$\varphi(n) := \max\{\psi(n, F) \mid F \text{ formula}\}$$

to grow linearly or quadraticly in the input n.

Today we know that it is possible to reduce the set of all satisfiable propositional formulas, which is NP-complete, to Gödel's problem.[24] This could make it surprising that Gödel was inclined to view such a solution as being "quite within the realm of possibilities" since (i) $\varphi(n) \geq k \cdot n$ is "the only estimate obtainable by generalizing the proof of the unsolvability of the Entscheidungsproblem", and (ii) $\varphi(n) \sim k \cdot n$ (or $\varphi(n) \sim k \cdot n^2$) "just means that the number of steps when compared to pure trial and error can be reduced from N to $\log N$ (or $\log N^2$)".

Furthermore, Gödel gave an interesting interpretation of the implications of this solution. If Gödel's question could be answered in the positive, then there would be a feasible algorithm yielding, among the formulas of first order logic, those which are provable in a reasonable sense of the word. Gödel thought this outcome to mean that the thinking of a human mathematician dealing with questions with a yes-or-no answer could be replaced, with just the exception of the formulation of the axioms for mathematics, by a machine. He thought this conclusion to refine what one can conclude under the unsolvability of the general decision problem, which would allow to consider any Turing machine as unfit to substitute the activity of a human mathematician. This conclusion would not in fact apply for a suitable choice of the input of the machine solving the problem above.

In fact, under the kind of pragmatic approach to human reasoning we've roughly described before, it is unreasonable to look for unbounded proofs. Then, if n is chosen sufficiently large, the activity of a human mathematician is in this sense reducible to that of a machine, provided Gödel's problem is solvable. In other words, although the human mind may be intensionally different from a Turing machine, it would be in fact equivalent to it for any practically feasible purpose.

[23] See [6, pp. 373–376].

[24] Such a reduction, as perfomed by Stephen A. Cook (see [2, p. v]), goes as follows: given a propositional formula A with atoms $p_1, \ldots, p_n$, and given a new monadic predicate symbol Q and individual variables $x_1, \ldots, x_n$, let $A' \equiv A[p_i := Q(x_i)]$ ($1 \leq i \leq n$) and $B \equiv \exists x \exists y(Q(x) \wedge \neg Q(y)) \rightarrow \exists x_1 \ldots \exists x_n A'$. Then $A \in SAT$ iff B is provable and it has a polinomially bounded proof.

5 Some General Comments

As a moral of our survey, one could naturally be brought to draw a balance of the interaction between a philosophically motivated problem and a mathematical treatement of it.

It could in fact be argued that the conclusion that follows from the complexity argument we've just accounted for, is basically the same of the one that was obtained by means of the modal argument. However, there is something in the approach to the problem of the previous section which is worth emphasizing as a distinguishing feature.

Gödel in fact thought the argument based on the unprovability of consistency and its disjunctive conclusion, to be the best possible result for a mathematical account of the situation concerning mind and machines. By that he plausibly meant that any further clarification, and even a solution of that disjunction should be sought by means of conceptual analysis. This would cause the whole problem to be subject to the unrigorous state of the philosophical investigation. The $P = NP$ argument may then represent, so to say, a mathematical way to such a solution which makes it unnecessary any further speculation of a philosophical nature.

This may have an unexpected consequence as to a comparison between Gödel and Turing's analysis. While in fact it seems natural to oppose the philosophical committment of the one to the neutrality of the other, Gödel's approach would end up with a conclusion which is even more radical than the one which can be attached to Turing's efforts. Indeed, one may find it justified to conclude that not only pure philosophy can be dispensed with in case of certain philosophically-committed scientific goals, but that it can be entirely substituted by making an appropriate use of certain results belonging to the most rigorous forms of a rational investigation. This claim would be further supported by the fact that the basic conceptual problem admits a straightforward and clear formulation (as attested by statement ($*$) above), so that it can be subject to a mathematical treatement.[25] However, this general view is contrasted by the difficulty in reaching a definitive solution by treating the corresponding problem mathematically.

In the case of Gödel's former argument, this difficulty is attested by the disjunctive conclusion that was obtained under a plausible definition of the concepts involved. As to Gödel's complexity argument instead, the weak aspect of it coincides with its depending on the solution of the $P = NP$ problem which, contrary to Gödel's expectations, would be nowadays recognized as unlikely.

[25]Notice that this conclusion is only apparently opposed to the independent meaning of the philosophical investigation. Indeed, it could be suitably reformulated in such a way that it would recall some classical approaches to philosophy (the best and most obvious example of which would be Leibniz's view), which were based on the belief that its behaving according to the canons of deductive reasoning should be one of its required features.

References

[1] Bringsjord, S., Arkoudas, K.: The modal argument for hypercomputing minds. Theoretical Computer Science **317** (2004) 167–190

[2] Clote, P., Krajíček, J.: Arithmetic, Proof Theory, and Computational Complexity. Oxford University Press (1993)

[3] Gödel, K.: Some Remarks on the Undecidability Results. *Collected Works vol. II: Publications 1938-1974.* Oxford University Press (1990) 305–306

[4] Gödel, K.: Undecidable Diophantine Propositions. *Collected Works vol. III: Unpublished Essays and Lectures.* Oxford University Press (1995) 157–175

[5] Gödel, K.: Some Basic Theorems on the Foundations of Mathematics and Their Implications. *Collected Works vol. III: Unpublished Essays and Lectures.* Oxford University Press (1995) 304–323

[6] Gödel, K.: *Collected Works vol. V: Correspondence H-Z.* Oxford University Press (2003)

[7] Hodges, A.: Alan Turing: The Enigma. Burnett Books, Simon and Schuster (1983)

[8] Lindstöm, P.: Penrose's New Argument. Journal of Philosophical Logic **30** (2001) 241–250

[9] Lindstöm, P.: Remarks on Penrose's "New Argument". To appear in the Journal of Philosophical Logic (online version already available at web page of this Journal)

[10] Lucas, J. R.: Minds, Machines and Gödel. Philosophy **36** (1961) 112–127

[11] Lucas, J. R.: Minds, Machines and Gödel: A Retrospect. P. Millican and A. Clark (eds) *Machines and Thought: The Legacy of Alan Turing.* Mind Association Occasional Series, Clarendon Press (1996) 103–124

[12] Penrose, R.: Shadows of the Mind. Oxford University Press (1994)

[13] Penrose, R.: Commentaries on 'Shadows of the Mind'. Psyche **2** (1995)

[14] Pudlák, P.: A Note on the Applicability of the Incompleteness Theorems to Human Mind. Annals of Pure and Applied Logic **96** (1999) 335–342

[15] Shapiro, S.: Mechanism, Truth, and Penrose's new argument. Journal of Philosophical Logic **32** (2003) 19–42

[16] Turing, A. M.: Systems of Logic Based on Ordinals. Proc. of the London Math. Soc. **45** (1939) 161–228

[17] Turing, A. M.: Computing Machinery and Intelligence. Mind **59** (1950) 433–460

[18] Turing, A. M.: Lecture to the London Mathematical Society on 20 February 1947. D. C. Ince (ed.) *The Collected Works of A. M. Turing, vol. II.* North Holland (1992) 87–105

[19] Turing, A. M.: Intelligent Machinery. D. C. Ince (ed.) *The Collected Works of A. M. Turing, vol. II.* North Holland (1992) 107–127

[20] Wang, H.: *From Mathematics to Philosophy.* Routledge and Kegan Paul (1974)