Is Human Mind Fully Algorithmic? Remarks on Kurt Gödel's Incompleteness Theorems

Mircea Dumitru(✉)

University of Bucharest and Romanian Academy,
Bdul M. Kogalniceanu nr. 36-46, 050107 Bucharest, Romania
mircea.dumitru@unibuc.ro

Abstract. In this paper I shall address an issue in philosophy of mind related to philosophy of mathematics, or more specifically to the nature of mathematical knowledge and reasoning. The issue concerns whether the human mind is fully algorithmic. I shall develop my answer against the background which is created by Kurt Gödel's celebrated incompleteness theorems. In what follows: (i) I shall first sketch the main programs and responses to the mind-body problem in philosophy of mind; (ii) then, I shall provide an informal overview of the two Gödelian incompleteness theorems; (iii) finally, I shall present and comment upon some of the main views advocated by Gödel about minds and machines, mind and matter, and the contrast between Turing machines and the so-called Gödel minds. In the process, Gödel's very unorthodox and unfashionable views against computabilism, neuralism, physicalism, psychoneural parallelism, and even against the underlying philosophical presuppositions of the Turing machines will emerge. Shocking as they, understandably, are, as compared to the standard psychological and philosophical orthodoxy underlying the received computabilistic views on mind, Gödel's own views are worth exploring and they fully deserve our undivided philosophical attention. Gödel is, after all, the founding father and one of the essential inspiring sources for the whole domain and range of topics that I address in my paper.

Philosophy of mind is thriving nowadays. The field has been developed extensively and intensively receiving all sorts of input from other connected fields, notably from computer science and cognitive science. The complexity and vitality of the domain is reflected by the vast literature which ramifies in various sub-fields and directions of research in which one tackles a batch of interrelated topics: the ontological problem (the so-called mind-body problem), the semantic problem, the epistemological problem, the methodological problem, artificial intelligence, and problems of neuroscience.

In my paper I shall deal with an issue in philosophy of mind related to philosophy of mathematics, or more specifically to the nature of mathematical knowledge and reasoning. The issue is whether or not human mind and intelligent consciousness is fully algorithmic. I shall develop my answer against the background which is created by Kurt Gödel's celebrated incompleteness theorems. In what follows: (i) I shall first sketch the main programs and responses

A. Beckmann et al. (Eds.): CiE 2015, LNCS 9136, pp. 23–33, 2015.
DOI: 10.1007/978-3-319-20028-6_3

to the mind-body problem in philosophy of mind; (ii) then, I shall provide an informal overview of the two Gödelian incompleteness theorems; (iii) finally, I shall present and comment upon some of the main views advocated by Gödel about minds and machines, mind and matter, and the contrast between Turing machines and the so-called Gödel minds. In the process, Gödel's very unorthodox and unfashionable views against computabilism, neuralism, physicalism, psychoneural parallelism, and even against the underlying philosophical presuppositions of the Turing machines will emerge. Shocking as they, understandably, are as compared to the standard psychological and philosophical orthodoxy underlying the received computabilistic views on mind, Gödel's own views are worth exploring and they fully deserve our undivided philosophical attention. Gödel is, after all, the founding father and one of the essential inspiring sources for the whole domain and range of topics that I address in my paper. And even if we do not, and perhaps cannot, take everything that he thought on those issues on board, one can still have a lot to learn from how he framed the questions and what he had to say about those fascinating issues concerning the nature and the functioning of our (mathematical) mind.

1 Sketch of the Main Programs and Responses to the Mind-Body Problem in Philosophy of Mind

So, let us first canvass the metaphysics of mind. A very useful resource for this topic is [1], whom I basically follow for the systematization of the main philosophical responses to the mind-body problem What we aim at clarifying here is the problem of the nature of the mind's states and processes. More specifically, the questions that we raise are: where do mental states and processes occur, and how are they correlated to the physical world? Is my consciousness going to survive my physical decay after I am dead? Is it possible that a purely physical system (a computer) be built in such a way that it can have conscious experience with qualia? Where do minds come from? What are they?

The reasoned answers to those difficult issues are theory and methodological driven. They are dependent upon the particular theory of mind that we may favor, which is based on its explicative and predictive power, and also on its coherence and simplicity. The main theories that have been advocated in the philosophy of mind are:

Dualism. The essence of all forms of dualism, such as substance dualism and property dualism, is that the nature of the mental resides in a nonphysical entity, which escapes the domain of physics, neurophysiology and computer science. Dualism, nowadays, undergoes a sort of paradoxical fate. By far the most popular and traditional philosophical perspective on mind, akin to the position advocated by various major religions on the relation between mind, soul, and body, dualism is almost completely rejected by professional philosophers today.

Philosophical behaviorism. This has been a tremendously influential conception in the metaphysics of mind for several decades in the XX-th century.

The rise of cognitivism in linguistics and psychology led to the demise of this once powerful position and critical tool against traditional speculative metaphysics. As such, philosophical behaviorism is not a theory about the essence of mental states per se; it is, rather, a kind of analysis of the language in which we talk about our mental states. Thus, sentences about various mental episodes, such as emotions, sensations, beliefs, desires, wants, etc. are not about would-be inner occurrences of mental events, but instead, they are abbreviated ways of speaking about actual and possible behavior. Therefore, any sentence about a mental state or process can be rephrased in a longer sentence about behavior.

Reductive materialism (Identity theory). The main claim of this form of materialist theory is that mental states are (identical with) physical states of the brain. More specifically, each type (or token, in weaker versions of this theory) of mental states or processes is numerically identical to some type (token) of physical state or process which takes place in the brain or in the central nervous system.

Functionalism. This doctrine, which is the prevalent view on mind today, says that the characteristic feature of any type of mental state is the set of causal relations it bears to the input coming from the environment, to other types of mental states, and to the output of our behavior. A mental state plays a causal role, and that mental state is defined through its network of causal roles. According to functionalism, as opposed to philosophical behaviorism, reference to mental states cannot be eliminated; and in order to define such a type of mental state, one has to refer to a number of other mental states with which that state is causally connected. Functionalism acknowledges the reality of mental states which should be studied systematically. It follows that psychology should be an autonomous science from, and not reducible to, the physical sciences (physics, biology, neurophysiology). Psychology is a science in its own right with its own irreducible laws, and its own domain.

Eliminative materialism. This is a profoundly skeptical view on the mind. It casts doubts upon the concepts and explanations of folk psychology (which explains the intelligent actions of human beings in terms of the causal powers of propositional attitude ascriptions, such as belief, desire, hope, etc.). Eliminative materialism also goes against reductive materialism, since part of that reduction program is to achieve a one-to-one correspondence between the mental states and processes acknowledged by folk psychology and some neuro-physiological processes that occur in the brain. This intended reduction cannot be done. And the reason is not lack of ingenuity from the part of the theorists. The reason is the non-existence of such things as mental states, processes or attitudes which are posited by this common-sense psychological framework which, in its turn, is hypothesized by folk psychology through an inference to the best explanation. Therefore, one key reason for this reduction being impossible resides in the fact that the common-sense psychological framework is fraught with some fatal problems: it is literally false, and consequently, it is also a misleading conception of what determines causally our behavior and mental activity. Through scientific

education, it is expected that, and it is hoped that, gradually one can get rid of this false representation about our own psychology, based on propositional attitude ascriptions. This framework will be eliminated by future neuroscientific discovery. This is the motivation for the name of the conception, viz. "eliminative materialism".

A general idea that emerges from the various responses to the mind-body problem, which is essentially a leit-motif of the ongoing dialectics running through those questions and answers, is that what best explains mental states and processes is a computational paradigm of the mind. Most theorists argue that the mind is a sort of computation on symbols and representational mental contents. This computational paradigm will offer a coherent answer to the hard problem of integrating two distinct views on human beings: the causal view, which underlies the explanation of the bio-chemical complex structures in which human beings qua biological entities consist, with the intentionality of the mental representation view of human beings qua rational and socio-cultural decision-making agents. The hope is that functionalism will solve this integration problem. The most influential version of current functionalism considers a computational theory of mind to be the best available explanation of human behavior via the causal role of mental states to mediate, explicitly in computational terms, between the environmental input and the behavioral output. We shall see in a moment that Gödel rejects both this view on mind and its presuppositions.

2 A Short Informal Overview of the Two Gödelian Incompleteness Theorems

What Gödel's First Incompleteness Theorem shows is that any consistent formal axiom system or deductive system T, which is sound (i.e. proves only true sentences) and powerful enough to express elementary arithmetic, is bound to be incomplete because a sentence, that we shall call G_T, can be true according to the interpretation of that formal system T, but cannot be derived as a theorem in that system.

Thus, Gödel shows that the common idea, according to which arithmetical truth equals proof within a formal deductive system, is wrong. Gödel was able to prove this following a series of ingenious steps of ([4], p. 1–7). First, he constructed a sentence G_T, in the language of arithmetic (via the technique of Gödel-numbering), which represents the meta-mathematical sentence: "The sentence G_T is not provable in the system T". That is, G_T says of itself that it is unprovable in T. It follows that G_T is true if and only if (iff) G_T cannot be proved in T. Let's suppose further that T is sound. If G_T were provable in T then G_T would be false, and hence unprovable in T, since T is sound and it can only prove true sentences. So, up to this point, if G_T were provable, then it could not be proven. Therefore G_T is not provable after all in T meaning that G_T is true. Suppose now that G_T were not provable. Then G_T is true and, of course, its negation, $\sim G_T$, is false. But T is sound and it proves only true sentences. Thus, T cannot prove $\sim G_T$ either. So, there is a true sentence, G_T, which says

of itself that it is not provable in a system T, and neither that sentence G_T, nor its negation $\sim G_T$ is provable in T. Hence the sentence G_T is undecidable by the means of the system T and, assuming that T is sound, the system T is incomplete. Adding G_T to the system T does not solve the issue because, according to the same method, a new sentence G'_T can be constructed in such a way as to be able to say of itself that it is not provable in $T + G_T$, while being true, and while neither G'_T nor its negation $\sim G'_T$ being provable in $T + G_T$.

Thus far Gödel has shown that, since G_T is true and unprovable in T, the axioms of the system T are incomplete. Summing up this part of the proof, that culminates in Gödel's First Incompleteness Theorem, Nagel & Newman in ([3], p. 67) cogently argue that "we cannot deduce all arithmetical truths from the axioms. Moreover, Gödel established that arithmetic is essentially incomplete: even if additional axioms were assumed so that the true formula G_T could be formally derived from the augmented set, another true but formally undecidable formula could be constructed."

In the Second Incompleteness Theorem, Gödel shows how to construct an arithmetical statement A that has the meta-mathematical content: "Arithmetic is consistent". He goes one to prove that the sentence "$A \longrightarrow G_T$" is formally provable; however, since G_T itself is not provable, Gödel shows that A is not provable either. What follows from this is the Second Incompleteness Theorem which establishes the fact "that the consistency of arithmetic cannot be established by an argument that can be represented in the [very same] formal arithmetical calculus" ([3], p. 67).

Do all these Gödelian ground-braking meta-mathematical results have any philosophical significance? And if so, what would that significance be? From among many reactions and comments that Gödel's Incompleteness Theorems have prompted[1], I shall take a look at Gödel's own philosophical views correlated with his own results, and make some comments on three issues, concerning (a) Gödel's view on minds, machines and computabilism, (b) Gödel's view on mind, matter, physicalism, and psycho-physical parallelism, and (c) Gödel vs. Turing, i.e. Gödel's view on Turing Machines and on Gödel Minds.

3 Comments upon Some of the Main Views Advocated by Gödel

This section of the paper is based on [6], which is an extremely reach source for Gödel's philosophical views.

3.1 Gödel About Minds and Machines

Gödel had a strong conviction that neither computabilism, i.e. the view that the brain and the mind work essentially like a computer, nor neuralism, i.e. the view that the brain is a sufficient explanans for mental phenomena is right, and

[1] For some of the reactions see [2,5,6].

consequently he argued vigorously against and rejected both views. Hao Wang [6] tells us that Gödel was preoccupied with the problem of whether computabilism was a complete explanation of mental processes, "that is, the issue of whether all thinking is computational - with special emphasis on mathematical thinking. Gödel's main concern was to demonstrate that not all mathematical thinking is computational" ([6], p. 183).

In one if its several formulations, Gödel's Second Incompleteness Theorem states something that is relevant to the mathematical capacity of the human mind, namely that if a reasonably strong theorem-proving computer or program is sound and consistent, then it cannot prove the truth that expresses its own consistency. According to [6], Gödel drew a relevant conclusion from this concerning the human mind: "6.1.1 The human mind is incapable of formulating (or mechanizing) all its mathematical intuitions. That is, if it has succeeded in formulating some of them, this very fact yields new intuitive knowledge, for example the consistency of this formalism. This fact may be called the "incompletability" of mathematics. On the other hand, on the basis of what has been proved so far, it remains possible that there may exist (and even be empirically discoverable) a theorem-proving machine which in fact is equivalent to mathematical intuition, but cannot be proved to be so, nor even be proved to yield only correct theorems of finitary number theory" ([6], p. 184–185).

Hao Wang, again, tells us that Gödel was very attached to some ideas about creation in mathematics and the algorithmic nature of human mind and mathematical thought ([6], p. 186). Those ideas are relevant for the implications of his theorem, as one can see from the following remark made by Gödel: "6.1.8 My incompleteness theorem makes it likely that mind is not mechanical, or else mind cannot understand its own mechanism. If my result is taken together with the rationalistic attitude which Hilbert had and which was not refuted by my results, then [we can infer] the sharp result that mind is not mechanical. This is so, because, if the mind were a machine, there would, contrary to this rationalistic attitude, exist number-theoretic questions undecidable for the human mind" ([6], p. 186–187).

The upshot of all those remarks is Gödel's strong conviction that human mind, through its intuitive powers and creativity, is superior over computers, and that the partaking of individual minds to the collective experience of the human species gives a whole new range of possibilities, which allows the human mind and spirit to surpass the power of computing machines. Here are some of Gödel's thoughts in this regard:

"6.1.19 The brain is a computing machine connected with a spirit."
"6.1.21 Consciousness is connected with one unity. A machine is composed of parts."
"6.1.23 By *mind* I mean an individual mind of unlimited life span[2]. This is still different from the collective mind of the species. Imagine a person engaged in solving a whole set of problems: this is close to reality; people constantly introduce new axioms." ([6], p. 189).

[2] Gödel believed the human soul is immortal, that science will prove that fact one day. His philosophical hero was Leibniz.

This fragment tellingly shows Gödel's trust that eventually we can prove mind's superiority over computers, because of its creativity and power to give new forceful ideas and insights:
"6.1.24 It would be a result of great interest to prove that the shortest decision procedure requires a long time to decide comparatively short propositions. More specifically, it may be possible to prove: For every decidable system and every decision procedure for it, there exists some proposition of length less than 200 whose shortest proof is longer that 10^{20}. Such a result would actually mean that computers cannot replace the human mind, which can give short proofs by giving a new idea." ([6], p. 189).

3.2 Gödel About Mind and Matter

The making of the distinction between mind and matter imposes upon us the metaphysical idea that they are distinct from each other, a doctrine which commits us to some form of dualism (see above). The difficulty to tie the two together (causally or otherwise), once we separated them essentially, has been notorious since Descartes' work. Gödel has his own way of framing this celebrated metaphysical issue, namely by asking whether "the brain suffices for the explanation of all mental phenomena". Gödel rephrases the question in a more precise, quantitative fashion, raising the issue of whether there are enough brain operations that represent the mental operations in such a manner that the correspondence between physical brain and mental operations is one-to-one or even many-to-one.

Scientific and philosophical orthodoxy argues that such a correlation exists, a view which is known as *psychoneural parallelism*. If, further on, one makes the physicalist assumption, which is quite common today, that all neural operations are physical operations of a special kind, the view turns into *psychophisical parallelism*.

Gödel's own argument is that there is mind which is separate from brain (matter). Gödel, contrary to the whole scientific establishment and current orthodoxy, refutes both psychoneural and psychophysical parallelism. His remarks with regard to this topic are very daring, and surely shocking for many of us. Thus, says Gödel: "6.2.1 Parallelism is a prejudice of our time. 6.2.2 Parallelism will be disproved scientifically (perhaps by the fact that there aren't enough nerve cells to perform the observable operations of the mind)" ([6], p. 190).

Gödel is quick to recognize that not all prejudices are necessarily false. A prejudice is a widely shared belief whose strength is not backed by solid pieces of evidence. Why do we hold so strongly to the parallelism prejudice? We do it, because we are impressed by the power of science and technology, often leading us to uncritically accepting scientism. Gödel makes the further extraordinary remark that the philosophical point of the parallelism in the aforementioned 6.2.2 is not only a philosophical prejudice, but also a scientific and empirical stance that will be disproved. Gödel emphasizes this idea whenever he feels that it is important to make more room for it in conceptual space. He refers to this notion in the following passages: "6.2.3 It is a logical possibility that the

existence of mind [separated from matter] is an empirically decidable question. This possibility is not a conjecture. [...] there is an empirical question behind it. 6.2.4 Logic deals with more general concepts; monadology, which contains general laws of biology, is more specific. The limits of science: Is it possible that all mind activities [...] are brain activities? There can be a factual answer to this question. Saying no to thinking as a property of a specific nature calls for saying no also to elementary particles. Matter and mind are two different things. 6.2.5 The mere possibility that there may not be enough nerve cells to perform the function of the mind introduces an empirical component into the problem of mind and matter" ([6], p. 191).

Gödel puts a lot of emphasis on the important and difficult metaphysical issue of the relation between mind and matter, considering it central to philosophical inquiry and critical to understanding philosophy's importance to science. Thus, one can read the following remark made by Gödel in conversation with Hao Wang: "6.2.6 Many so-called philosophical problems are scientific problems, only not yet treated by scientists. One example is whether mind is separate from matter. Such problems should be discussed by philosophers before scientists are ready to discuss them, so that philosophy has as one of its functions to guide scientific research. Another function of philosophy is to study what the meaning of the world is" ([6], p. 191).

Gödel is very interested in clarifying the issue of the parallelism between mind and matter, clearly stating his stance: "6.2.9 Mind is separate from matter: it is a separate object ..." Moreover, he boldly conjectures, completely going against the grain of the scientific establishment, that science itself will eventually refute this prejudice of the psychoneural parallelism: "6.2.11 [...] I believe that mechanism in biology[3] is a prejudice of our time which will be disproved. In this case, one disproof, in my opinion, will consist in a mathematical theorem to the effect that the formation within geological times of a human body by the laws of physics (or any other laws of a similar nature), starting from a random distribution of the elementary particles and the field, is as unlikely as the separation by chance of the atmosphere into its components" ([6], p. 192).

And a last remark in this regard. The remark shows Gödel's conviction that: (a) the brain is a physical object, (b) the mind (or the spirit) is a separate entity from the brain, and (c) the brain, as a normal physical object, functions the way it does just because it is connected to a mind: "6.2.14 Even if the finite brain cannot store an infinite amount of information, the spirit may be able to. The brain is a computing machine connected with a spirit. If the brain is taken to be

[3] By 'mechanism in biology' Hao Wang says that Gödel meant Darwinism, "which he apparently sees as a set of algorithmic laws (of evolution). Even though he seems to believe that the brain - and presumably also the human body - functions like a computer [...], he appears to be saying here that the human body is so complex that the laws of physics and evolution are insufficient to account for its formation within the commonly estimated period of time" ([6], p. 192).

physical and as [to be] a digital computer, from quantum mechanics [it follows that] there are then only a finite number of states. Only by connecting it [the brain] to a spirit might it work in some other way" ([6], p. 193).

3.3 Turing Machines vs. the So-Called Gödel Minds

Gödel thought profoundly of the nature of algorithms, and of the formalization of logical systems. Consequently, he was very interested in the ground-braking work of Alan Turing, holding Turing's work in very high esteem. Gödel came to believe that his own incompleteness theorems hit upon an important aspect of the limits of formalization only after Turing developed his analysis, which Gödel fully endorsed, of the concept of mechanical (or computational) procedures, the so-called Turing machine. Moreover, Gödel was satisfied with the fact that Turing machines provide evidence for the thesis that sharp concepts really exist, and that human minds can perceive them clearly. Nevertheless, Gödel spotted a problem in Turing's argument of the adequacy of his analysis of algorithms, namely a fallacious proof of the conclusion that minds and machines are equivalent ([6], p. 194).

Gödel's position is made clear through the following remark: "6.3.5 Attempted proofs for the equivalence of minds and machines are fallacious. One example is Turing's alleged proof that every mental procedure for producing an infinite series of integers is equivalent to a mechanical procedure" ([6], p. 197). Gödel explains why he considers the proof attempted by Turing to be fallacious: "6.3.6 Turing gives an argument which is supposed to show that mental procedures cannot carry farther than mechanical procedures. However, this argument is inconclusive, because it depends on the supposition that a finite mind is capable of only a finite number of distinguishable states" ([6], p. 197).

Gödel rejects the supposition that mind (spirit) is matter; he says: "6.3.7 It is a prejudice of our time that (1) there is no mind separate from matter; indeed, (1) will be disproved scientifically" ([6], p. 198). He, then, continues to interpret and reconstruct Turing's argument, and finds it valid, only after certain presuppositions are guaranteed and accepted: "6.3.8 It is very likely that (2) the brain functions basically like a digital computer. 6.3.9 It is practically certain that (2') the physical laws, in their observable consequences, have a finite limit of precision. 6.3.10 If we accept (1), together with either (2) or (2'), then Turing's argument becomes valid" ([6], p. 198).

It is hard for us today not to accept all these presuppositions just because we are accustomed, or indeed perhaps prejudiced, to thinking of the brain and the mind as being two aspects of the same thing. However, Gödel did not consider the matter to be so: "6.3.11 If (i) a finite mind is capable only of a finite number of distinguishable states, then (ii) mental procedures cannot carry any farther than mechanical procedures. 6.3.12 Turing's argument (iii) for the condition (i) is his idea which centers on the following sentence: *We will also suppose that the number of states of mind which need be taken into account is finite. The reasons for this are of the same character as those which restricted the number of symbols.*

If we admit an infinity of states of mind, some of them will be 'arbitrarily close' and will be confused." ([6], p. 198).

Gödel is happy with the inference from (i) to (ii). He believes, though, that (i) can be inferred from (iii) only if some additional assumptions are forthcoming. And since Gödel does not accept that brain is equivalent to mind, he goes on to reject both (i) and (ii).

At the end of the day, Gödel's refutation of mental computerism, his deeply held conviction that mind can carry farther than machines, is based on the idea that "*6.3.13 Mind, in its use, is not static, but constantly developing.*" When we focus, introspectively, on the stream of our consciousness, we are struck by the fact that the mental states and their succession do not enjoy the sharpness and clarity of the states of Turing machines. Wang comments the following concerning Gödel's idea: "... we develop over time, both individually and collectively; and so, for instance, what appeared to be complex becomes simple, and we understand things we did not understand before. Here again, we feel that the process of development is somewhat indefinite and not mechanical" ([6], p. 200).

Do we have a proof that minds can carry farther than computers, and that they are not fully mechanical? We do not. However, Gödel promotes a dynamic and developing kind of vision of mind that is both telling and credible: "6.3.14 Although at each stage of the mind's development the number of its possible states is finite, there is no reason why this number should not converge to infinity in the course of its development" ([6], p. 200).

A mechanical brain connected to a creative, ever-evolving, developing, and non-mechanical mind (spirit) is a set-up that goes beyond the individualism of the atomic, and isolated minds. Thus, brains and minds can create thoughts in a manner which reflects that minds can carry farther than brains and computers, eventually indicating mind's superiority over computers.

Acknowledgements. I want to thank Dr. Victor Mitrana, University of Bucharest, for commenting upon an earlier version of the paper which contributed to the improvement of the arguments, and for introducing the text in LaTeX. I am grateful to Dr. Daniela Dumitru, Bucharest University of Economic Studies, for stimulating discussions about cognitivism and computerism. I also want to thank Miss Ioana Andrada Dumitru, PhD student at Johns Hopkins University, for making comments and stylistic suggestions which I happily accepted and which improved the clarity and the readability of the paper.

References

1. Churchland, P.M.: Matter and Consciousness. The MIT Press, Cambridge (1992)
2. Hintikka, J.: On Gödel. Wadsworth, Belmont (2000)
3. Nagel, E., Newman, J.R.: Gödel's Proof. Routledge, London (1958)

4. Smith, P.: An Introduction to Gödel's Theorems, 2nd edn. Cambridge, New York (2013)
5. Tieszen, R.: After Gödel. Platonism and Rationalism in Mathematics and Logic. Oxford, UK (2011)
6. Wang, H.: A Logical Journey. From Gödel to Philosophy. The MIT Press, Cambridge (1996)