Must all Mathematicians be Platonists?
A case on Penrose's use of Gödel

Matthias Eichhoff, *Department of Political Science, Free University of Berlin*
Gerald Weber, *Department of Mathematics and Computer Science, Free University of Berlin*

Abstract

Roger Penrose has received broad attention for his arguments with the conclusion "the human mind is nonalgorithmic". From this position he concludes that a Quantum Gravitational Theory must be nonalgorithmic. In his writings Penrose discusses Gödel's famous incompleteness sentence and his epistemological position he calls Mathematical Platonism. In our article we reconstruct the implicit logical structure of Penrose's argumentation with the following results:

First, we show that his conclusion "the human mind is nonalgorithmic" can be obtained if *both* Gödel's sentence *and* Mathematical Platonism are taken as premises.

Second, we show that Penrose originally derives his conclusion solely from Gödel's sentence, using a certain interpretation that differs from Gödel's own.

Third, it is shown that in both cases the practice of mathematics would prescribe a certain epistemology, namely Mathematical Platonism.

We argue that Penrose did not recognize the constitutional, indispensable function of Mathematical Platonism for his considerations and the ensuing consequences for the epistemological state of his whole argument.

1. Introduction

Must all Mathematicians be Platonists? Is it true, as Roger Penrose states, that Kurt Gödel's incompleteness sentence forces us to a Platonic view on Mathematics? And what is the function of Penrose's Platonic interpretation of Gödel's sentence for his nonalgorithmic-mind-thesis? These are the questions discussed in this article in order to show the different epistemological standpoints of Penrose and Gödel, and their consequences for the Philosophy of Mathematics.

Roger Penrose, Rouse Ball Professor of Mathematics at the University of Oxford, discusses in Part I of his book *Shadows of the Mind* Gödel's famous incompleteness theorem and the epistemological position called Mathematical Platonism, deriving as central conclusion "the human mind is nonalgorithmic". From this controversal position he concludes that a Quantum Gravitational Theory must be nonalgorithmic, and he elaborates a scientific program in Part II of this book. The book is a sequel to the earlier *The Emperor's New Mind*. These positions were subject to broad discussions, for example in *The Large, the Small and the Human Mind* and several other publications.

In our article we reconstruct the until now hidden logical structure implicit in Penrose's argumentation with the following results:

First, we show in Chapter 2 that his conclusion "the human mind is nonalgorithmic" can be obtained if *both* Gödel's sentence *and* Mathematical Platonism are taken as premises.

Second, we explain in Chapter 3 that Penrose originally derives his conclusion solely from Gödel's sentence, using a certain philosophical interpretation that differs from Gödel's own (see Chapter 4).

Third, it is shown that in both methods of deduction, the practice of mathematics would prescribe a certain epistemology, namely Mathematical Platonism.

Chapter 5 provides an assessment on the power of Penrose's argumentation. We argue that Penrose did not recognize the constitutional, indispensable function of Mathematical Platonism for his considerations and the ensuing consequences for the epistemological state of his whole argument.[1]

2. The logical structure of Penrose's argument reconstructed

We start our considerations by presenting a reconstruction of Penrose's non-algorithmic-mind argument, which yields the same conclusion, but proceeds differently as his original argumentation. We present our reconstruction first, because it introduces the notions involved, and the formalism needed for our considerations. Afterwards this reconstructed argument will serve as a frame of reference for the explanation of Penrose's original argument.
Our reconstruction of Penrose's argument has the form of a syllogism, i.e. a logical inference from two premises, a major premise and a minor premise, to a resulting conclusion. Our choice to present our reconstruction as a syllogism is a result and not a prerequisite of our analysis. This choice enables us to discuss the logical structure with a minimum of formalism.

The *major premise* of Penrose's argument is Gödel's famous incompleteness theorem, one of the most important contributions to logic. This theorem initiated a development in which the connections between formal logic and the theory of computability were settled, and several reformulations of the original theorem were found. Penrose himself emphasizes that this theorem plays a central role in his argumentation. In a later remark to his original paper, Gödel pointed out that the definition of computability found by Turing should also be used to develop the definition of formal systems, and that he would have used this definition himself in his incompleteness sentence, if it would have been available at this time. We follow here Gödel's advice and choose a reformulation that incorporates the notion of computability or algorithmicity due to Turing and inspirations from a contribution of Rosser and the truth set theory of Tarski. This reformulation deals with sets of sentences that are consistent and complete. If we call such a set that contains furthermore all sentences that are derivable in a sufficiently strong axiomatic system of mathematics[2] a *truth set*, then Gödel's sentence says:
No truth set is algorithmic (or recursively enumerable, or semi-decidable, as it is often called in theory).
We want to introduce some abbreviations, so if a set is a truth set we simply say it is T, and if a set is algorithmic, we say it is A. Now we can write Gödel's sentence as:

Major Premise: No T is A.

In search for an *minor premise* we look at Penrose's epistemological position about the ability of the human mind to recognize mathematical truth, which he himself calls “Platonism” (for example in [4], p. 50), and which we will call Mathematical Platonism because of its strong connections to mathematics. Penrose claims, that there is one objective truth set of which human minds are able to ascertain whether a given sentence is in this set (i.e. is true) or not.
If we call such a set of sentences, for which membership is ascertainable for human minds, a *recognizable* set, and abbreviate the property to be a recognizable set as R, then the position of Mathematical Platonism can be written as:

Minor Premise: Some T are R. [3]

It is important to be aware that the claim of existence of some T and some R is implicitly contained in the formulation of the minor premise. Now if we take Gödel's sentence as major premise and Mathematical Platonism as minor premise, then both allow a way of reasoning which is called *modus ferison* in syllogistic, and this leads to the conclusion:

Conclusion: Some R are not A.

This means *expressis verbis*: Some sets of sentences (i.e. at least one set of sentences) that are recognizable to the human mind are nonalgorithmic. But this is nothing else than Penrose's conclusion: The human mind is nonalgorithmic. Thus we have verified the derivability of his conclusion from these two premises.

3. Penrose's original deduction

On the basis of our reconstruction in the last chapter we now turn to Penrose's original argument. The same three statements, which we have reconstructed as major premise, minor premise and conclusion, are considered, but their mutual logical dependence is established in quite another way.

Penrose himself claims that his conclusion can be derived from Gödel's sentence without further premises. In the Preface of [4] he writes (page vi): "Central to the arguments in Part I, is the famous theorem of Gödel, and a very thorough examination of the relevant implications of Gödel's sentence is provided. [...] The conclusions are that conscious thinking must indeed involve ingredients that cannot be even simulated adequately by mere computation; [...] Accordingly the mind must indeed be something that cannot be described in any kind of computational terms."

In the same book, Penrose presents this thought even clearer. In [4], I 1.15, p. 50, he writes: "If, as I believe, the Gödel argument is consequently forcing us into an acceptance of some form of viewpoint *C* (*the viewpoint that awareness is evoked by a physical, but nonalgorithmic brain process, the authors*) then we shall also have to come to terms with some other of its implications. We shall find ourselves driven towards a *Platonic* viewpoint of things."

Penrose has been interpreted in this way by supporters of his arguments, too. Malcolm Longair writes in the foreword to [5]: "Such a computer cannot discover mathematical theorems in the way that human mathematicians do. This surprising conclusion is derived from a variant of what is called Gödel's Theorem. [...] Because of the central importance of this result for his general argument, he *(i.e. Penrose, the authors)* devoted over half of Shadows of the Mind to showing that his interpretation of Gödel's Theorem was watertight."

4. Some consequences of Penrose's position for the Philosophy of Mathematics

In both the original and the reconstructed argumentation an existential quantifier for recognizable and nonalgorithmic sets has to be introduced to derive Penrose's non-algorithmic-mind thesis.[4]

But where does it stem from? On which place in the deduction has it been introduced?

In our reconstruction of Penrose's argumentation the existential quantifier enters the stage along with the minor premise "Some T are R". It stems, inherently plausible, from Mathematical Platonism.

However, in Penrose's original deduction it follows directly from Gödel's sentence. This latter implies that his interpretation of Gödel's sentence contains an existential quantifier for recognizable and nonalgorithmic sets. In other words, Penrose infers a Platonistic view for his nonalgorithmic mind thesis from a Platonistic interpretation of Gödel's sentence. It is interesting to consider if Gödel himself deduced Mathematical Platonism from his incompleteness sentence. He did not.

Though Gödel's philosophical standpoint was that mathematical objects have an existence in a sense comparable to the existence of physical bodies, he pointed out that his ontological position was not decisive for the question of truth in mathematics. He wrote in [9], p. 272: "However, the question of the objective existence of objects of mathematical intuition (which, incidentally, is an exact replica of the objective existence of the outer world) is not decisive for the problem under discussion here. The mere psychological fact of the existence of an intuition which is sufficiently clear to produce the axioms of set theory and an open series of extension to them suffices to the question of the truth or falsity of propositions like Cantors Continuum Hypothesis.”

Gödel's philosophical standpoint is quite different from Penrose's position, especially regarding the exactness of our intuition of the mathematical axioms.

He writes in [8], p. 213: “The analogy between mathematics and a natural science is enlarged upon by Russel also in another respect (in one of his earlier writings). He compares the axioms of logic and mathematics with the laws of nature and logical evidence with sense perception, so that the axioms need not necessarily be evident in themselves, but their justification lies (exactly as in physics) in the fact, that they make it possible for these ‘sense perceptions’ to be deduced; which of course would not exclude that they also have a kind of intrinsic plausibility similar to that in physics. I think that (provided ‘evidence’ is understood in a sufficiently strict sense) this view has been largely justified by subsequent developments, and it is to be expected that it will be still more so in the future. [...] Of course, under these circumstances mathematics may lose a good deal of its ‘absolute certainty;’ but, under the influence of the modern criticism of the foundations, this has already happened to a large extent.”

In contrast to this, in Penrose's argument it is absolutely essential, that the human perception of mathematical thought should be in principle exact. The human mind recognizes truth if and only if truth can be ascertained and not only be guessed.

We have seen that Penrose's position implies that mathematics prescribe a certain philosophy, namely Mathematical Platonism. The reader should notice that our objective was only to show the immanent consequences of Penrose's argument. Neither leaves his interpretation room for other epistemological positions than Mathematical Platonism, nor does he discard any epistemological concept at all in a, say, turn to Pragmatism. Hence, whoever supports Penrose's claim must also in fact come to terms with its consequences and necessarily accept Mathematical Platonism as Philosophy of Mathematics. According to this position, everybody having another epistemological attitude with respect to mathematics must either unconsciously violate his position while doing mathematics, or he must have failed to understand the mathematical method.

5. Conclusion

Though there are many variants of Mathematical Platonism, it can be roughly said that this realistic viewpoint states that the objects of mathematics are eternal and independent of the human mind. Two of the most challenging problems of Mathematical Platonism are how we can know of these objects, and how they are tied to natural sciences in applied mathematics. Penrose's introduction of his, even though pretty inscrutable, epistemological notion of “insight”, and his research in physics, shall bridge this gap. According to Penrose, noncomputability is the common property of both minds and mathematics, which could make it more plausible to understand, in his theory, how human beings know of Platonic objects; and a non-algorithmic Quantum Graviational Theory could even provide a unified theory of science. Contemporary Philosophy of Mathematics must indeed cope with natural sciences, but how can such a project work if its premises are already contaminated with dogmatic, non-empirical notions? If Penrose had introduced noncomputability as a transcendental “somewhere in there” concept for both minds and mathematics, and not as a Platonistic metaphysical one with mathematical objects “somewhere out there”, there had been hope for a clearer and more distinct understanding of what is human. Unfortunately for Penrose, his manoeuvre leads to a more speculative epistemological state of his nonalgorithmic mind thesis than to a foundation of a Platonic view of the natural sciences.

Notes

[1] We thank Raúl Rojas for his remarks and advice in finishing this article.

[2] A set of sentences is defined to be consistent and complete if it contains for each sentence either the sentence or its negation. An example of a sufficiently strong theory is the theory of computation, as it was introduced by Turing and earlier by Church, with some usual axiomatization. This theory contains statements, whether computing processes halt or not. For this theory the Gödel sentence becomes almost the same as Turing's result, that the halting problem is not computable. A truth set for this theory is the set of all true sentences of either the affirmative form “Machine m starting from configuration x halts” or the negative form “Machine m starting from configuration x does not halt.” While the subset of this truth set containing all true affirmative sentences is computable, namely semi-decidable, the subset containing all true negative sentences is not computable.

[3] Penrose would surely make this sentence even stronger as “Exactly one T is R”, but our argument holds even in the version given here.

[4] Penrose says in [4], p380, “there is no requirement that all halting problems are accessible to human understanding and insight,” that means he does not claim, that all true mathematical sentences are recognizable. Hovever, this does not answer the question, where the existential qantifier for the remaining nonalgorithmic set of recognizable true sentences stems from. But on the other hand, it makes his position most unclear: Is there a further, unrecognizable kind of true mathematical sentences?
At this point it should be mentioned that Penrose means the whole of recognizable true sentences to be not only nonalgorithmic, but even not diagonizable. That means, that a Gödel-type argument does not lead out of the recognizable true sentences. This is surely the consequent formulation of Mathematical Platonism. However, this again does not answer the question about the origin of the necessary existence quantor.

Bibliography:

[1] K. Gödel (1931), *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatshefte für Mathematik und Physik,**38**, 173-98.

[2] R. Penrose (1989), *The Emperor's New Mind*, Oxford Univ. Press, Oxford.

[3] R. Penrose et al. (1990)., *Precis of The Emperor's New Mind and open Peer Commentary*, The Behavioral and Brain Sciences,**13,4** pp 643-705.

[4] R. Penrose (1994), *Shadows of the Mind*, Oxford Univ. Press, Oxford.

[5] R. Penrose (1997), *The Large, the Small and the Human Mind*, Cambridge Univ. Press, Cambridge.

[6] Rosser, J. B. (1936). *Extensions of some theorems of Gödel and Church*, Jour. Symbolic Logic,**1**, 87-91.

[7] A. M. Turing, (1937) *On computable numbers with an application to the Entscheidungsproblem*, Proc. Lond. Math. Soc. (ser. 2), **42**, 230-65; a correction **43**, 544-6.

[8] K. Gödel (1944), *Russel's Mathematical Logic*, in Paul A. Schilpp, ed., The Philosophy of Bertrand Russell, The Library of Living Philosophers, Evanston, Ill. (New York: The Tudor Publishing Company, 1944), pp. 125-52;
reprinted in: P. Benacerraf and H. Putnam, eds.,*Philosophy of Mathematics*, Prentice Hall, 1964, pp. 211-32.

[9] K. Gödel (1947), *What is Cantor's Continuum Problem?*, The American Mathematical Monthly, **54**, 515-25;
revised and reprinted in: P. Benacerraf and H. Putnam, eds.,*Philosophy of Mathematics*, Prentice Hall, 1964, pp. 258-73.