Mathematics and the Mind

by Edward Nelson
Department of Mathematics
Princeton University
http://www.math.princeton.edu./∼nelson/papers.html

The claim of J. R. Lucas [7, 8], expounded by Roger Penrose in the section entitled "The non-algorithmic nature of mathematical insight" in *The Emperor's New Mind* [9, Chap. 10], can be phrased as follows:

> Gödel's theorem shows that there is no algorithm producing all statements of Arithmetic that mathematicians can see to be true. Therefore consciousness acts non-algorithmically.

I shall analyze this claim and then present some contrasting thoughts.

What are statements of Arithmetic? Here there is no dispute. They are expressions formed from symbols

$$\forall \quad \exists \quad \neg \quad \& \quad \vee \quad \rightarrow \quad = \quad 0 \quad \mathrm{S} \quad + \quad \cdot \quad (\quad) \quad x \quad y \quad z \quad \ldots$$

according to certain rules. For example,

$$\forall x \exists y (\exists a \, x + a = y \ \& \ \forall m \forall n (m \cdot n = y \ \vee \ m \cdot n = y + \mathrm{SS0} \ \rightarrow \ m = \mathrm{S0} \ \vee \ n = \mathrm{S0}))$$

is a statement of Arithmetic.

What does it mean for a statement of Arithmetic to be true? It depends on whom you ask. Students of the foundations of mathematics fall into three schools: *Platonists*, such as Gödel, *intuitionists*, such as Brouwer, and *formalists*, such as Hilbert. I shall first explain the Platonic notion of truth in our example.

A lexicon helps:

$\forall$	for all
$\exists$	there exists
$\neg$	not
&	and
$\vee$	or
$\rightarrow$	implies
$=$	equals
0	zero
S	successor
$+$	plus
$\cdot$	times

Think of the variables $x\ y\ z\ \ldots$ as ranging over the natural numbers

$$0, \quad S0, \quad SS0, \quad SSS0, \quad \ldots$$

(i.e., 0, 1, 2, 3, ...). Then Platonists say that our formula is true in case for all numbers x, there exists a larger number y (larger since $x+a=y$) such that both y and $y+2$ are primes (if either one can be factored as $m \cdot n$, then one of the factors m or n is 1). This is a famous unsolved problem of Arithmetic, the twin primes problem. So far, no mathematician can prove it or disprove it.

Intuitionists demand an effective method for constructing the twin primes greater than any given x before assenting to the truth of the formula. Our example is a very simple statement for which it is easy to acquire an intuition, but a truth definition for statements of Arithmetic must apply also to statements containing thousands of symbols and not corresponding to any notions for which we have built up an intuition. The Platonic truth definition for a statement $\exists x_1 \forall y_1 \exists x_2 \forall y_2 \ldots$ reads: there exists a number x_1 such that for all numbers y_1 there exists a number x_2 such that for all numbers y_2 ... This notion of interleaved infinite searches and verifications makes sense only to Platonists; to intuitionists and formalists the idea lacks coherence.

Kleene [5, §82] has made a deep analysis of the intuitionistic truth notion in terms of recursive realizability. He shows that intuitionistic truth is not simply a restriction of the Platonic truth notion but actually conflicts with it. That is, he exhibits a statement of Arithmetic that is Platonically true and intuitionistically false, and another that is Platonically false and intuitionistically true.

Formalists find Platonic truth for statements of Arithmetic both incoherent and irrelevant to the practice of mathematics. A solution, positive or negative, of the twin primes problem would be a major event in mathematics, but no solution will depend on philosophical truth notions.

Penrose, in footnote 2 to the last chapter of [9], says "As to the very dogmatic Gödel-immune formalist who claims not even to recognize that there *is* such a thing as mathematical truth, I shall simply ignore him, since he apparently does not possess the truth-divining quality that the discussion is all about!" He explicitly ignores formalists and tacitly ignores intuitionists, thereby dismissing two of the century's greatest mathematicians and students of foundations.

Platonic truth is conceived to be an infinite, abstract structure that assigns to each statement of Arithmetic a truth value, true or false. Platonists believe in the existence, though not in the physical world, of this ideal object. Intuitionists believe in a different ideal object, intuitionistic truth, and formalists believe in neither. These differing philosophical views are unproblematic for mathematics—which is concerned with proof (syntax) rather than truth (semantics)—but belief in Platonic truth is an essential part of the Lucas argument. It is a truism that mathematics has many applications to science. But an application to science of the philosophy of mathematics (or, more accurately, of one philosophy of mathematics) would indeed be a novelty.

What is Gödel's theorem? As already described, statements of Arithmetic are simply symbols combined according to certain rules. Arithmetic is a powerful language and the syntactical rules for forming statements can be expressed, or encoded, within Arithmetic.

To obtain a theory of Arithmetic, call it $\mathcal{A}$, a mathematician selects certain statements of Arithmetic as axioms and postulates certain rules of deduction. A statement of Arithmetic is provable in $\mathcal{A}$ in case it follows from the axioms by the rules of deduction. This is a syntactical notion on which there is no dispute, and it too can be expressed in Arithmetic. But the choice of axioms and rules of deduction varies according to one's semantic notion of truth and one's personal aesthetics. It is essential that $\mathcal{A}$ be consistent, since otherwise every statement of Arithmetic could be proved in $\mathcal{A}$, and $\mathcal{A}$ would be of no interest. One much studied theory $\mathcal{A}$ has come to be known as Peano Arithmetic.

The liar paradox, famous from antiquity, can be formulated as

this statement is not true.

The liar paradox plays no role in Gödel's theorem. The notion of Platonic truth is semantic, not syntactical, and cannot be expressed within Arithmetic. (This was proved by Tarski [10].) Gödel does something different. He constructs a statement of Arithmetic, call it G, expressing

this statement is not provable in $\mathcal{A}$.

Using many abbreviations, the construction of G takes a fair number of pages. The modern form of Gödel's first incompleteness theorem [3] is

if $\mathcal{A}$ is consistent, then G is not provable in $\mathcal{A}$.

This is a theorem, and there is no dispute about it. Notice that Gödel's theorem says nothing whatever about consciousness or the mind.

The Lucas argument, endorsed by Penrose, can now be phrased as follows:

> Mathematicians, by their consciousness, see that $\mathcal{A}$ is consistent and therefore know that G is not provable in $\mathcal{A}$, and therefore know that G is true. Thus mathematicians, by their consciousness, have insight—going beyond proof—into mathematical truth. Hence consciousness produces truth by non-algorithmic means.

The claim is not that mathematicians have direct insight into the extremely complicated statement G; rather it is that they follow the proof of Gödel's theorem, have direct insight into the hypothesis that $\mathcal{A}$ is consistent, and so conclude that G is true even though unprovable in $\mathcal{A}$.

But can mathematicians see, without proof, that certain statements of mathematics are true? A look at the history of mathematics shows that this belief has invariably been a delusion. Mathematicians thought they could see that cubic equations could only be solved geometrically, using conic sections—until del Ferro, Tartaglia, and Cardano solved them algebraically; they thought they could see that Euclid's parallel postulate was true—until Lobachevsky and Bolyai constructed non-Euclidean geometry.

Today we are beginning (some of us) to learn caution. The fact of the matter is that serious students of the foundations of mathematics are divided about the status of the hypothesis that Peano Arithmetic is consistent. Platonists claim to see it directly;

intuitionists can be convinced by another, much less cited, theorem of Gödel [4] combined with their claimed insight into their own notion of truth; some formalists regard it as a genuinely open question. And the problem becomes more acute when, as required by the Lucas argument, we study general theories $\mathcal{A}$ and general algorithms.

What is an algorithm? Although important specific algorithms date from antiquity, and the word itself comes from the name of the ninth century algebraist al-Khwarizmi, the general nature of an algorithm was not elucidated until the 1930s, primarily by Turing [11]. Today we can express the notion succinctly by saying that an algorithm is a computer program that, for any input, eventually halts, producing an output. This is another Platonic notion, presupposing an infinite search over all inputs and all numbers (the number of steps required for the computer to halt, if it does halt). Turing proved that there is no algorithm to decide whether any given computer program is an algorithm (undecidability of the halting problem).

A closely related result is that there is no algorithm that decides whether any given theory $\mathcal{A}$ is consistent. Lucas and Penrose state that mathematicians have direct insight into the consistency of some theories $\mathcal{A}$, but they do not specify on which $\mathcal{A}$ this faculty of insight operates correctly.

Now we have at hand all the ingredients of the Lucas-Penrose argument. They postulate belief in an infinite ideal object known as Platonic truth and they postulate a faculty of consciousness to perceive by direct insight the consistency of certain unspecified theories $\mathcal{A}$. They dismiss those mathematicians who hold different views and conclude that they have uncovered a general and fundamental feature of consciousness.

The similarity to religious belief, with appeal to authority, is inescapable. Here is a close analogue of the Lucas argument: Catholics know that when the Pope speaks *ex cathedra* on matters of faith or morals, he speaks the truth; they then know those truths without proof; therefore consciousness acts non-algorithmically. (I hasten to say, sincerely, that I offer this analogy simply to clarify my view of the Lucas argument. I respect Catholics and Catholic beliefs, and I respect Platonists.)

I conclude that Gödel's theorem, and the general Platonic notion of an algorithm, are simply irrelevant to the study of the mind. But can the study of algorithms offer anything of relevance to a science of consciousness? I am not a student of consciousness, but it does seem safe to say that natural selection has produced in us conscious awareness that can perform certain computations very rapidly and efficiently. Therefore it is plausible that a mathematical study, not of Platonic algorithms in general but of rapid and efficient algorithms, might have some relevance to a science of consciousness.

Such a study is a relatively new field, more developed in computer science departments than in mathematics departments. The usual way of making precise the notion of a rapid algorithm is to say that the computer program always halts after a number of steps bounded by a given polynomial in the length of the input. There is still a Platonic element here: the bound is simply said to exist. But Bellantoni and Cook [1, 2] and Leivant [6] have produced a purely syntactical notion and demonstrated that it is equivalent to the notion of a polynomial-time algorithm, This is a surprising result. It is in sharp contrast to the Turing result on the undecidability of the halting problem for general algorithms. There are no arithmetic hypotheses in their characterization; rather, their idea expresses the

Aristotelian notion of an object coming into being, as opposed to the Platonic notion of a pre-existing ideal object.

The study of the similarities and differences between polynomial-time algorithms and the functioning of the mind is a concrete matter, not just philosophical speculation, and it will repay the serious attention it is receiving.

Here is a vision that I expect to become a reality early in the twenty-first century. Mathematicians write articles with precisely structured statements of theorems and proofs. The proofs refer to the database of previously proved theorems, a database that lives in a decentralized fashion on the Web. The proof is automatically checked for correctness by a polynomial-time algorithm, and if correct becomes a part of the database. Such an article requires no refereeing for correctness. The system is used interactively to help find proofs. For mathematics that deals with concrete, syntactical objects—numbers, graphs, anything that can be encoded by a string of bits—and for proofs that are constructive, the system automatically constructs from the proof a program that produces the existentially quantified objects as functions of the relevant universally quantified objects, together with a worst-case bound on time and memory requirements. This will be a polynomial-time bound if the recursions in the proof satisfy the Bellantoni-Cook-Leivant constraint. Such a program requires no debugging since it is a by-product of a proof that the objects have the properties stated in the theorem whose constructive proof has been verified.

The difficulty in searching for a mathematical proof is the exponentially growing sea of possibilities that opens up as one possible step among many follows another possible step. What will be of interest to practicing mathematicians in such a system is the interactive exploitation of the different search strategies of conscious human minds and computers. What will be of interest to students of Artificial Intelligence, and possibly to a science of consciousness, is that the system will provide a laboratory for testing whether human search abilities—which are a marvel and a mystery—can be simulated by machine algorithms. (My own hunch is that it will be centuries, if ever, before computers—or, more likely, robots with a body and some experience of the world—can begin to do this.) This is a modest question that concerns the functioning of the mind in one area, rather than the fundamental nature of consciousness, but I expect it to be more fruitful than speculations about Platonic truth and algorithms.

References

[1] Stephen Bellantoni and Stephen Cook, "A new recursion-theoretic characterization of the polytime functions", *Computational Complexity*, **2**, 97-110, 1992.

[2] Stephen Bellantoni, "Predicative Recursion and Computational Complexity", *Technical Report 264/92*, Department of Computer Science, University of Toronto, 1992. `ftp://ftp.cs.toronto.edu/pub/reports/theory/cs-92-264.ps.Z`

[3] Kurt Gödel, "Über formal unentscheidbare Sätze der Principia Mathematica und verwanter Systeme I", *Monatshefte für Mathematik und Physik*, **37**, 173-198, 1931.

[4] —, "Zur intuitionistischen Aritmetik und Zahlentheorie", *Ergebnisse eines math. Koll.*, **4**, 34-38, 1933.

[5] Stephen Cole Kleene, "Introduction to Metamathematics", North-Holland, 1952.

[6]Daniel Leivant, "Ramified recurrence and computational complexity I: Word recurrence and poly-time", in P. Clote and J. Remmel (eds.), *Feasible Mathematics II*, 320-343, 1994.

[7] J. R. Lucas, "Minds, machines and Gödel", *Philosophy*, **36**, 102-124, 1961.

[8] —, "The Godelian Argument", *Truth*, **2**, 1988. `http://www.clm.org/truth/2truth08.html`

[9] Roger Penrose, "The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics", Oxford University Press, 1989.

[10] Alfred Tarski, "Der Wahrheitsbegriff in den formalizierten Sprachen", *Studia philosophica*, **1**, 261-405, 1936. Translated from the Polish original of 1933 by L. Blaustein.

[11] Alan Turing, "On computable numbers with an application to the Entscheidungsproblem", *Proc. London Math. Soc.*, **42**, 230-275, 1936. "A correction", *ibid.*, **43**, 544-546, 1937.

nelson@math.princeton.edu