



planetmath.org

Math for the people, by the people.

Penrose’s first Gödelian argument

|                  |                               |
|------------------|-------------------------------|
| Canonical name   | PenrosesFirstGodelianArgument |
| Date of creation | 2013-03-22 17:05:45           |
| Last modified on | 2013-03-22 17:05:45           |
| Owner            | dankomed (17058)              |
| Last modified by | dankomed (17058)              |
| Numerical id     | 23                            |
| Author           | dankomed (17058)              |
| Entry type       | Topic                         |
| Classification   | msc 03D80                     |

The application of Gödel's theorems to outside metamathematics, notably the philosophy of mind, was initiated by Gödel himself. He had a strong philosophical bent towards Platonism, which also motivated his mathematical discoveries (Saint-Andre, 1998). Gödel first thought that his theorems established the superiority of mind over machine (Wang, 1996a; 1996b). Later, he came to a less decisive, conditional view: if machine can equal mind, the fact that it does cannot be proved (Bojadziev, 1997).

"either...the human mind ( within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable diophantine problems."  
 – Gödel, 1951\*, p. 310

The importance of Gödel's theorem for understanding of human mind was revived by Lucas (1961) and was brought in the scope of brain scientists recently by Roger Penrose (1989, 1994).

In 1989 Roger Penrose suggested that human mind is nonalgorithmic (noncomputable) and more powerful than any formal system and this follows from Gödel's first theorem. Indeed his argument aimed to show that we (humans) can unmistakably find a true statement  $G(\check{g})$  for given formal system  $F$ , which the formal system  $F$  cannot prove, but which we (humans) know is true. Thus stated the statement for mind noncomputability follows from statement which is stronger than Gödel's first theorem, and due to numerous errors and ambiguities in the exposition it is not clear whether Roger Penrose in his book "*The Emperor's New Mind*" claims to have proved mind noncomputability from result stronger than Gödel's first theorem, or proves it directly from Gödel's first theorem. That is why below we provide concise but not equivalent formulations of Penrose's main thesis. Indeed it seems that Penrose himself makes no difference between these two.

Penrose's first Gödelian argument (formulation 1, mind noncomputability is derived from statement stronger than Gödel's first theorem): Human mathematician sees the truthness of the Gödelian sentence  $G(\check{g})$  of given formal system  $F$ , therefore human mind is nonalgorithmic (noncomputable).

Penrose's first Gödelian argument (formulation 2, mind noncomputability is derived directly from Gödel's first theorem): Human mathematician understands (proves) Gödel's first theorem, therefore human mind is nonalgorithmic (noncomputable).

Both of these arguments are provably erroneous, and as stated are false. The correct full text of Gödel's first theorem that can be proven by every sufficiently strong formal system is the following.

Gödel's first theorem: for every sufficiently strong to capture basic arithmetic formal system  $F$  which is recursively axiomatizable (i.e. there is an algorithm for determining for each string of the underlying language whether it is an axiom or not) in which all valid sentences are representable by finite strings of symbols there exists true sentence  $G(\check{g})$ , which is undecidable within  $F$  if  $F$  is consistent. (Note:  $\check{g}$  is the numeral of the Gödel number  $g$  of the formula  $G(\check{g})$ )

Here we should note that if the formal system is inconsistent it will be able to prove any statement including the Gödel's first theorem. If the formal system is consistent and sufficiently strong it will also be able to prove Gödel's first theorem, and this was shown by Gödel's himself in 1930.

Gödel's first theorem is both human understandable and it is a theorem in any sufficiently strong finite size formal system  $F$  that captures basic arithmetic. Thus computers also prove Gödel's first theorem.

Penrose's argument (formulation 1) insisting that we directly see the truthness of  $G(\check{g})$  in  $F$  is false because we have to be able to directly decide the consistency of arbitrary finite size formal system  $F$ , which is clearly not the case. Therefore what Penrose has overlooked is that human when given a formal system  $F$  must be able to decide nonalgorithmically and unmistakably whether  $F$  is consistent or not, in order to determine whether  $G(\check{g})$  is true or not. Clearly  $G(\check{g})$  is false for inconsistent  $F$ ! Since no human can decide nonalgorithmically and unmistakably whether arbitrary  $F$  is consistent or not then Penrose's first Gödelian argument (formulation 1) is false, and the human mathematician cannot see directly that  $G(\check{g})$  is true, as  $G(\check{g})$  might be false for inconsistent  $F$ .

Penrose's argument (formulation 2) is modification that derives mind non-computability directly from Gödel's first theorem in a fashion "*the human mathematician can "see" that  $G(\check{g})$  is true for consistent  $F$  however the consistent  $F$  cannot prove  $G(\check{g})$* ". It should be noted however that this latter claim is not strong enough for one to establish superiority of mind over formal systems (algorithms) and does not prove noncomputability of mind, because the human does not "see" the truthness of  $G(\check{g})$ , the human derives (proves)

it given the premise for consistent  $F$ . Thus Penrose puts the human and the formal system in not equivalent positions: the human knows explicitly that  $F$  is consistent, while the  $F$  is given no such access on grounds that consistent formal system cannot know its own consistency as implied by Gödel's second theorem. This inequivalence of the initial "given" (not "seen"!) knowledge with which the human and the formal system start leads to erroneous final conclusions in the book "*The Emperor's New Mind*", and makes it impossible for the reader to understand whether Penrose makes distinction between the presented above formulation 1 and formulation 2 of his argument. The text of Gödel's first theorem is provable within the formal system  $F$  and it says exactly that " $G(\check{g})$  is true for consistent  $F$ ". Therefore no matter what the ontology of our mind is, there is nothing novel about the Gödel's first theorem and about the Gödel's statement  $G(\check{g})$  that the human knows but the artificial intellect does not prove.

#### References

1. Bojadziev D (1997) <http://nl.ijs.si/~damjan/g-m-c.html> Mind versus Gödel. In: M. Gams, M. Paprzycki and X. Wu (eds.), *Mind Versus Computer*, IOS Press, pp. 202-210.
2. Crossley JN, Brickhill C, Ash C, Stillwell J, Williams N (1972) *What is mathematical logic?* Oxford University Press.
3. Detlovs V, Podnieks K (2006) <http://www.ltn.lv/~podnieks/mlog/ml.htm> Introduction to Mathematical Logic. Hypertextbook for students in mathematical logic
4. Gödel K (1931) Über formal unentscheidbare Sätze der Principia und verwandter Systeme I. *Monatshefte für Mathematik und Physik* **38**: 173-198.
5. Gödel K [1951\*] *Collected Works, Vol. III. Unpublished Essays and Lectures*. Oxford University Press, 1995.
6. Lucas JR (1961) Minds, Machines and Gödel. *Philosophy* **36**: 112-127.
7. Mendelson E (1997) *Introduction to Mathematical Logic*, 4th ed. London: Chapman & Hall.
8. Penrose R (1989) *The Emperor's New Mind: Concerning Computers, Minds, and The Laws of Physics*. Oxford University Press.
9. Penrose R (1994) *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford University Press.
10. Podnieks K (2006) <http://www.ltn.lv/~podnieks/gt.html> What is Mathematics? Gödel's Theorem and Around. Hypertextbook for students in mathematical logic
11. Saint-Andre P (1998) <http://www.saint-andre.com/thoughts/wang-godel.html> Objectiv Without Platonism: Hao Wang on Kurt Gödel

12. Wang H (1996a) *A Logical Journey: From Gödel to Philosophy*. Cambridge, Massachusetts, The MIT Press.
13. Wang H (1996b) <http://www.fortunecity.com/emachines/e11/86/moremind.html> Can Minds do more than Brains? In *How Things Are: A Science Tool-Kit for the Mind*, John Brockman and Katinka Matson (eds.)