



planetmath.org

Math for the people, by the people.

Penrose’s second Gödelian argument

Canonical name	PenrosesSecondGodelianArgument
Date of creation	2013-03-22 17:05:48
Last modified on	2013-03-22 17:05:48
Owner	dankomed (17058)
Last modified by	dankomed (17058)
Numerical id	11
Author	dankomed (17058)
Entry type	Topic
Classification	msc 03D80

In the book *"Shadows of the Mind"* published in 1994 Roger Penrose revisited his argument against the strong AI (artificial intelligence) thesis. In contrast with Penrose's first Gödelian argument which used Gödel's first theorem, in 1994 Penrose used modified argument based on Gödel's second theorem in an attempt to prove that human understanding is beyond computation. Besides of numerous errors in the technical exposition (Feferman, 1995), the argument only shows that strong AI thesis is necessarily false, however is not strong enough to establish superiority of human mind over formal systems (robot mathematicians).

Strong AI (artificial intelligence) thesis: the human mind is consistent formal algorithm (system).

The argument against strong AI is direct with the use of Gödel's second theorem. If we were consistent formal algorithm we wouldn't be able to prove our consistency.

Penrose's second Gödelian argument: strong AI is false, because if we were consistent formal algorithm we wouldn't be able to prove our consistency.

Some strong AI advocates erroneously tried to escape the argument by philosophical argument that we "believe" we are consistent formal algorithm, and we do not prove it. This is elementary logical fallacy. Simply our scientific theory of mind is based on axioms (postulates), which we accept to be true and we accept that these postulates do correctly describe the reality. Therefore in our scientific knowledge strong AI thesis is either theorem or axiom. What some AI advocates overlook is the fact that axioms are provable! Simply an axiom is provable by any other axiom, and this is basic postulate in mathematical logic. Indeed more generally - every true statement is derivable from any other statement, including the possibility that it is derivable (provable) from arbitrary false statement. This is indeed the first axiom of propositional calculus  $B \Rightarrow (C \Rightarrow B)$ .

Therefore Penrose's second Gödelian argument is correct and strong AI is false. One cannot save the strong AI thesis by playing with the meaning of the words "believe", "understand" or "know". Axioms exactly as theorems are provable, and the colloquial understanding that we "*believe in axioms but we cannot prove them*" is false. Axioms in any system  $F$  are provable within  $F$ .

Of course always remains the possibility for weaker AI thesis insisting that human mind is inconsistent algorithm. This however is not the strong AI thesis, and one must create and study new kind of paradoxical mathematics.

”We have seen that minds can behave both as consistent formal systems, and as inconsistent systems. Therefore, if we are to build a machine which can be compared fairly with a human mind it is only logical to expect the machine to behave both consistently and inconsistently, as well.” – Jeff Makey (1995)

Despite of the fact that Penrose’s second Gödelian argument is successful against strong AI thesis, Penrose’s original conviction that he has proved noncomputability of human mind appears to be flawed.

”Penrose’s arguments, if taken to their logical conclusion, show us not that the human mind is noncomputable, but that either the human mind is beyond all mathematics, or else we cannot be sure that it is consistent.” –Daryl McCullough (1995)

#### References

1. Detlovs V, Podnieks K (2006) <http://www.ltn.lv/~podnieks/mlog/ml.htm> *Introduction to Mathematical Logic*. Hypertextbook for students in mathematical logic
2. Feferman S (1995) <http://psyche.cs.monash.edu.au/v2/psyche-2-07-feferman.html> Penrose’s Gödelian Argument.
3. Makey (1995) <http://www.sdsc.edu/~jeff/Godel%5Fvs%5FAI.html> Gödel’s Incompleteness Theorem is Not an Obstacle to Artificial Intelligence.
4. McCullough D (1995) <http://psyche.cs.monash.edu.au/v2/psyche-2-04-mccullough.html> Humans Escape Gödel? A Review of *Shadows of the Mind* by Roger Penrose
5. Penrose R (1989) *The Emperor’s New Mind: Concerning Computers, Minds, and The Laws of Physics*. Oxford University Press.
6. Penrose R (1994) *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford University Press.
7. Penrose (1995) <http://psyche.cs.monash.edu.au/v2/psyche-2-23-penrose.html> Beyond the Doubting of a Shadow.
8. Podnieks K (2006) <http://www.ltn.lv/~podnieks/gt.html> *What is Mathematics? Gödel’s Theorem and Around*. Hypertextbook for students in mathematical logic