Appeared in: *Facta Philosophica* Vol. 4, No 1 (2002), 167-169.

McCall's Gödelian argument is invalid

Panu Raatikainen

Storrs McCall continues the tradition of Lucas and Penrose in an attempt to refute mechanism by appealing to Gödel's incompleteness theorem (McCall 2001). That is, McCall argues that Gödel's theorem "reveals a sharp dividing line between human and machine thinking". According to McCall, "[h]uman beings are familiar with the distinction between truth and theoremhood, but Turing machines cannot look beyond their own output". However, although McCall's argumentation is slightly more sophisticated than the earlier Gödelian anti-mechanist arguments, in the end it fails badly, as it is at odds with the logical facts.

McCall's reasoning differs from the earlier Gödelian arguments in his admission that the recognition of truth of Gödel sentence G_T for a theory T depends essentially on the unproved assumption that the theory T under consideration is consistent. But, so the argument continues, still human beings, but not Turing machines, can see that truth and provability part company. For, McCall notes, we can argue by cases:

> *Case 1.* T is consistent. G_T is unprovable, but true.
> *Case 2.* T is inconsistent. G_T is provable, but false.

Whichever alternative holds, McCall concludes, truth and provability fail to coincide. According to McCall, human beings can see this, but a Turing machine cannot.

The conclusion, however, is simply false. McCall does not seem to realize that e.g. in Peano Arithmetic PA one can formalize and prove Gödel's theorem for an arbitrary theory, however strong (McCall considers Zermelo-Fraenkel set theory ZFC, but one can take just any theory, e.g. ZFC + there exist supercompact cardinals, i.e. the sky is the limit, as long as the theory is effectively axiomatizable). More exactly, one can, in PA, formalize and weakly represent provability in any chosen theory T; let us denote such a provability predicate by $Prov_T(x)$. Gödel's self-referential trick (the diagonalization lemma) then provides a sentence G_T such that

(1) PA proves: $\neg Prov_T(G_T) \leftrightarrow G_T$.

By formalizing the proof of Gödel's (first) incompleteness theorem in PA, one obtains

(2) PA proves: $Cons(T) \rightarrow G_T$,

(where $Cons(T)$ is the sentence formalizing "T is consistent"), and by simple logic,

(3) PA proves: $Cons(T) \rightarrow [\neg Prov_T(G_T) \,\&\, G_T]$,

for an arbitrary theory T. And since one can construct a machine which enumerates all the theorems of PA, a machine can output all facts of this form. Hence a machine can make the distinction between sentences being true and being derivable in some theory.

At this point, one might object that the above formalized sentence (3) differs from McCall's original, informal conditional claim (T is consistent $\rightarrow$ G_T is unprovable but true), for the latter contains the notion of truth, which cannot, by Tarski's undefinability theorem, even be expressed in the language of arithmetic. The Gödel sentences, however, have a rather simple form; they are universal sentences (what logicians call Π_1 sentences), and for such a class of sentences, it is possible to define truth even in the language of arithmetic (indeed, by a Π_1 formula) (see e.g. Smorynski 1977, Hajek and Pudlak 1993); let us denote such a "partial truth-definition" by $Tr_1(x)$. One can then prove in PA, for an arbitrary formalized theory T, that:

(4) $Cons(T) \rightarrow [\neg Prov_T(G_T) \,\&\, Tr_1(G_T)]$.

Again, one can construct a machine which enumerates all facts of this sort. And such a machine can clearly "see" that truth and provability "part company".

One must thus conclude that McCall's new attempt to refute mechanism with the help of Gödel's theorem does not work. This does not, of course, mean that mechanism has to be true.

Panu Raatikainen
Department of Philosophy
University of Helsinki
panu.raatikainen@helsinki.fi

References

Hajek, Petr and Pudlak, Pavel. (1993) *Metamathematics of First-Order Arithmetic*, Springer-Verlag, Berlin.

McCall, Storrs. (2001) "On 'Seeing' the Truth of Gödel Sentence", *Facta Philosophica* 3: 25-29.

Smorynski, Graig. (1977) "Incompleteness theorems", in J. Barwise (ed.) *Handbook of Mathematical Logic*, North-Holland, Amsterdam, pp. 821-65.