

Three proofs of Undecidability of A_{TM}

(From G. Chaitin, "Gö del's Theorem and Information" *International J. Theor. Physics* 22 (1982), pp. 941-954.)

1. Define a function F where $F(N)$ is defined to be either 1 more than the value of the N th computable function applied to the natural number N , or zero if this value is undefined because the N th computer program does not halt on input N . F cannot be a computable function, for if program M calculated it, then one would have $F(M) = F(M) + 1$, which is impossible. But the only way that F can fail to be computable is because one cannot decide in general if the N th program ever halts when given input N .
2. This proof runs along the lines of Bertrand Russell's paradox of the set of all things that are not members of themselves. Consider programs for enumerating sets of natural numbers, and number these computer programs. Define a set of natural numbers consisting of the numbers of all programs which do not include their own number in their output set. This set of natural numbers cannot be recursively enumerable, for if it were listed by computer program N , one arrives at Russell's paradox of the barber in a small town who shaves all those and only those who do not shave themselves, and can neither shave himself nor avoid doing so. But the only way that this set can fail to be recursively enumerable is if it is impossible to decide whether or not a program ever outputs a specific natural number, and this is a variant of the halting problem.
3. Consider programs which take no input and which either produce a single natural number as output or loop forever without ever producing an output. Think of these programs as being written in binary notation, instead of as natural numbers as before. I now define a so-called Busy Beaver function: BB of N is the largest natural number output by any program less than N bits in size. The original Busy Beaver function measured program size in terms of the number of states in a Turing machine instead of using the more correct information-theoretic measure, bits. It is easy to see that BB of N grows more quickly than any computable function, and is therefore not computable, which as before implies that the halting problem is unsolvable.

The $3x + 1$ conjecture:

Define function f so that $f(n) = 3n+1$ if n is odd and $f(n) = n/2$ if n is even. The conjecture states that starting from any positive n , repeated iteration of this function produces 1.

On Gödel's Theorem

Gödel showed that within a rigidly logical system such as Russell and Whitehead had developed for arithmetic, propositions can be formulated that are undecidable or undemonstrable within the axioms of the system. That is, within the system, there exist certain clear-cut statements that can neither be proved or disproved. Hence one cannot, using the usual methods, be certain that the axioms of arithmetic will not lead to contradictions ... It appears to foredoom hope of mathematical certitude through use of the obvious methods. Perhaps doomed also, as a result, is the ideal of science - to devise a set of axioms from which all phenomena of the external world can be deduced.

Boyer, *History of Mathematics*.

All consistent axiomatic formulations of number theory include undecidable propositions

...

Gödel showed that provability is a weaker notion than truth, no matter what axiom system is involved ...

How can you figure out if you are sane? ... Once you begin to question your own sanity, you get trapped in an ever-tighter vortex of self-fulfilling prophecies, though the process is by no means inevitable. Everyone knows that the insane interpret the world via their own peculiarly consistent logic; how can you tell if your own logic is "peculiar" or not, given that you have only your own logic to judge itself? I don't see any answer. I am reminded of Gödel's second theorem, which implies that the only versions of formal number theory which assert their own consistency are inconsistent.

The other metaphorical analogue to Gödel's Theorem which I find provocative suggests that ultimately, we cannot understand our own mind/brains ... Just as we cannot see our faces with our own eyes, is it not inconceivable to expect that we cannot mirror our complete mental structures in the symbols which carry them out? All the limitative theorems of mathematics and the theory of computation suggest that once the ability to represent your own structure has reached a certain critical point, that is the kiss of death: it guarantees that you can never represent yourself totally.

D. Hofstadter, *Gödel, Escher, Bach*.

Remark: Many have argued (persuasively, I think) that Gödel's Incompleteness Theorem does not rule out strong AI (namely, computers that exhibit "human" reasoning) based on formal logic. Of course, this does not settle whether strong AI is possible.

The proof of Gödel's Incompleteness Theorem is so simple, and so sneaky, that it is almost embarrassing to relate. His basic procedure is as follows:

1. Someone introduces Gödel to a UTM, a machine that is supposed to be a Universal Truth Machine, capable of correctly answering any question at all.
2. Gödel asks for the program and the circuit design of the UTM. The program may be complicated, but it can only be finitely long. Call the program $P(\text{UTM})$ for Program of the Universal Truth Machine.
3. Smiling a little, Gödel writes out the following sentence: "The machine constructed on the basis of the program $P(\text{UTM})$ will never say that this sentence is true." Call this sentence G for Gödel. *Note that G is equivalent to: "UTM will never say G is true."*
4. Now Gödel laughs his high laugh and asks UTM whether G is true or not.
5. If UTM says G is true, then "UTM will never say G is true" is false. If "UTM will never say G is true" is false, then G is false (since $G = \text{"UTM will never say } G \text{ is true"}$). So if UTM says G is true, then G is in fact false, and UTM has made a false statement. So UTM will never say that G is true, since UTM makes only true statements.
6. We have established that UTM will never say G is true. So "UTM will never say G is true" is in fact a true statement. So G is true (since $G = \text{"UTM will never say } G \text{ is true"}$).
7. "I know a truth that UTM can never utter," Gödel says. "I know that G is true. UTM is not truly universal."

Think about it - it grows on you ...

With his great mathematical and logical genius, Gödel was able to find a way (for any given $P(\text{UTM})$) actually to write down a complicated polynomial equation that has a solution if and only if G is true. So G is not at all some vague or non-mathematical sentence. *G is a specific mathematical problem that we know the answer to, even though UTM does not!* So UTM does not, and cannot, embody a best and final theory of mathematics ...

Although this theorem can be stated and proved in a rigorously mathematical way, what it seems to say is that *rational thought can never penetrate to the final ultimate truth ...* But, paradoxically, to understand Gödel's proof is to find a sort of liberation. For many logic students, the final breakthrough to full understanding of the Incompleteness Theorem is practically a conversion experience.

Rucker, *Infinity and the Mind*.