

Comments on Gödel by Faustus from the Philosophy Forum

Here's a very dumbed down way to understand why Gödel is no threat at all to A.I..

All Gödel shows is that try as you might, you can't create any AI that is capable of proving the truth value of its own Gödel sentence. The Gödel sentence for that AI is the algorithm of the AI itself with a clever little marker up front. It's like having to evaluate the truth value of such sentences as "This sentence is false."

This is only a problem if you think that A.I. developers all along wanted to make perfect math proving machines. But they are concerned instead with perception, motor activity, memory, and other forms of reasoning, and don't demand perfection. After all, evolution made us through utterly mechanical processes, and we get by just fine being far short of ideal mathematical theorem provers. Furthermore, you could make a theorem-prover that was quite effective despite having various faults or even inconsistencies.

The theorem has no bearing whatsoever on the true aims of A.I., nor on mechanistic explanations for consciousness. But, for those interested in very narrow studies of rationality and mathematics, it may have some very interesting implications.

By Faustus (Brian Peterson)

I think these two quotes have some things in common:

Nosoul wrote:

EDIT: Still...Couldn't it still be unexpectedly important in unforeseen ways for an AI to be able to solve Goedel-problems? For all we know, yes, "solving Goedel problems" might be extremely trivial & specialized (sure seems like it to me); on the other hand, the ability to "solve" them (or comprehend them, at least) might be quite related to some large cluster of abilities & functionalities which, if an AI didn't have these, would banish the AI to a very significant functional lack of "human-like intelligence".

I mean, we just don't know, it seems to me.

muxol wrote:

It is not meant to be a charge against AI -- you've missed the point. It is a charge against mechanism which is quite different from AI. It is the view that minds are mechanical, and as such, minds are machines. The point is that if minds are capable of doing something that machines cannot, then we are obviously not machines at all regardless of the aims of AI.

The problem I have with both quotes is that I sense something about Gödel's Theorem isn't being appreciated, though Nosoul suspects it when he says it is "specialized".

If I'm remembering correctly, the theorem as written applied only to self-consistent provers of mathematical truths. It demonstrated that there is one thing any such prover cannot do. Then the Church-Turing thesis (?) proved that you could use Gödel's Theorem and apply it to any Turing machine, therefore by definition to any computer.

So, for our purposes the only Gödel problem to solve exists strictly for self consistent provers of mathematical truths that are Turing machines. And the problem has been defined so that it is impossible to solve. It's not a matter of having special abilities that another entity somewhere could lack, it's a matter of being the sort of thing to which the Theorem applies. If you are, you'll never solve it. If you aren't, it doesn't matter.

The young King Arthur is the one entity with “grace” enough to pull the sword out of the stone. He’s got what everyone else lacks. The illusion is that there is something parallel here. In this case, the sword is Gödel’s Theorem, and it is wondered if (or just asserted that) humans (or “minds” per Muxol) are the only beings that can pull it from the stone. Wrong, wrong, wrong!

If any given A.I. or just a mechanism is a self consistent prover of mathematical truths and is also at some level of description a Turing machine, then it could never “solve” its own Gödel sentence. Period, no negotiation. This sort of entity is the only one to which the Theorem applies.

If the same A.I. or mechanism were an inconsistent prover of mathematical truths and/or could not fit the description of any Turing machine, then Gödel has literally nothing to say about it. Here’s the most important point: this second kind of A.I./mechanism is all anyone who thinks minds are mechanisms has ever been arguing for. So the whole Gödel thing is moot.

It is exactly the same thing for human beings, who are bound to the same restrictions. If human beings are indeed self consistent provers of mathematical truths and also Turing machines, then we by definition have no capacity to pull the sword from the stone, either. But if we are not consistent and cannot be defined as Turing machines, then not only can we not pull the sword, it technically doesn’t even exist for us.

By Faustus (Brian Peterson)

Nomos wrote:

Gödel showed that human beings can know things that cannot be known by computation. Ergo, there is more to human beings than computation. I'm not sure there's any need to make the issue any more complicated.

Gödel showed absolutely no such thing—that’s the problem.

Nomos wrote:

The emphasis on G-sentences is a red herring.

Far from a red herring, it’s the key. Unless you understand the role of G-sentences, you don’t understand why this argument of Lucas fails.

You claimed that Gödel’s Theorem shows that human beings can know something without computation. Actually it doesn’t—Lucas and others have merely tried to harness it to their agenda of isolating the human mind from the natural world. But what the heck did you think that something we could “know” was? It is none other than that human being’s own personal G-sentence! The whole test, the whole point of invoking Gödel, was to show that we can do something a mechanism can’t. That something is the proving of our own Gödel sentences.

This idea is absurd in a variety of ways.

Absurdity Number One: I would like to see someone prove that in fact human beings are immune to the theorem. This gets asserted all over the place, but no one has ever made the case. I’d like to see it done—in fact, I demand to see it done! Unless this proof is made available, the entire argument rests upon the hunches and convictions of its supporters, and that just won’t do.

I really want to see an iron-clad demonstration that human beings can all solve their own Gödel sentences—and no cheating either! For instance, Muxol’s point to the effect that “a universal Turing machine can prove any Gödelian sentence G in some system $T+G$.” That is an example of what I would

call cheating. Of course the G sentence can be solved once you add to or change the original system. (The proper G-sentence at that point for our purposes would also have to include T, and then the Turing machine would once again be rendered helpless.) The point is that mechanisms and humans are equally effected (or not) by the Theorem. We do not possess magical abilities that make us immune. If you disagree, then prove rather than assume this is the case!

Absurdity Number Two: Considering that your G-sentence contains a sort of programming code describing your body and brain down to the lowest level, and sets out the rules for how this mechanism runs itself, that G-sentence must be pretty long. It has to describe the behavior of every atom and molecule in your body and what they do as you think (ie, it has to describe your entire mechanism in formal terms). That's a phenomenally long amount of code—far, far longer than anything you will ever run on a PC. I think it's quite likely that the Sun would be a red giant before you could even finish reading the thing. You would simply lack the memory and thinking power needed to even hope of trying to prove it true. And if you make use of some technology to help, then the mechanistic "rules" for every such tool—be it a computer or a pencil and paper—then have to become part of the G-sentence, or that would be the sort of cheating that Muxol described. Physically, it is literally impossible for a human being to even attempt to jump over Gödel's hurdle. (I know many are fond of "in principle" feats of heroism, but I'm not so fond of them myself.)

Absurdity Number Three: Even the attempt to claim human beings are immune to Gödel is itself subject to a sort of "liar's paradox" effect. Say you really think you could prove your own Gödel sentence and must be something more than a mechanism as a result. By definition, if you even have a Gödel sentence, then you must be some sort of Turing machine. Which means that you are conceding up front that you are some sort of mechanism even as you try to refute it.

Absurdity Number Four: I can't remember if anyone on this thread made the claim, but it has been made elsewhere: that our (mythical) ability to thwart the Theorem shows that mechanism cannot be the full explanation for our consciousness. Lucas assures us that we could still come up with valid, mechanistic models of the brain, but that some special ingredient would be missing—presumably what makes us conscious.

So let's completely drop the previous three absurdities with this whole thing and consider those human beings who are uneducated in mathematics, just stupid, perhaps retarded, or those who are children. This subset of humans is obviously not, in present form, ever going to be able to solve, let alone read, a Gödel sentence. Even Turing machines existing today perform far better than they at a variety of mathematical tasks. But they are conscious despite their inability to perform mathematical tasks, including the solving of their own Gödel sentences. Right?

Unless one is prepared to deny this, then it turns out Gödel's insights, in fact, have nothing to do with the relationship between mechanism and consciousness.

By Faustus (Brian Peterson)

Nomos wrote:

I don't care about Lucas. I'm talking about what Gödel proved, not what Lucas proved.

But Gödel's Theorem doesn't say anything about mechanism or the human mind, which seem to be the focus of your posts.

Nomos wrote:

Human reasoning is not immune to the theorem, but human beings are. This is the point.

Oh, humans are? Care to prove that rather than just asserting it?

Nomos wrote:

Human beings can decide questions that cannot be decided by computation. This is not news, but Gödel produced a mathematical proof.

Baloney. Gödel proved no such thing and didn't even attempt to do so—that attempt was only made by others much later. Again, I must ask that you back this stuff up with evidence and argument.

Nomos wrote:

But when we say 'something exists' we know that this is true. We cannot know this by a computational process.

Yet another blind-faith assertion. Can you prove it?

Nomos wrote:

I think we have different idea of what a G-sentence is. What do you mean by "your G-sentence contains a sort of programming code describing your body and brain down to the lowest level".

The Church-Turing thesis proved that Gödel's Theorem applied to computers (Turing machines). In this case, the lines of computer code making the Turing machine play the role that axioms did in the original Theorem. That's what a G-sentence becomes when applied to a Turing mechanism. Therefore, if you are saying that human beings are immune in any sense to the Theorem, you have to be saying that human beings can "solve" their own G-sentences, which means "solving" an incredibly huge bit of code. If you aren't saying this, then you aren't saying anything that the Theorem applies to.

Nomos wrote:

You seem to be arguing that human beings cannot know anything, which is not what most people conclude, and seems to be a self-defeating argument, since in this case you cannot know that you are right.

That's a hallucination on your part, as I said nothing of the sort.

Nomos wrote:

You are muddling up formal reasoning with knowing. They are not the same thing, as Gödel showed.

No, it's you who have muddled things up. Gödel didn't even attempt to say anything about knowing in the general sense of the term. The theorem only applies to a very technical aspect of mathematical proof and the limits of formal reasoning in that domain. If you aren't talking about formal proof, you aren't talking about Gödel.

Nomos wrote:

Of course our reasoning systems are subject to the incompleteness theorem. Nobody argues otherwise as far as I know.

Lucas argued otherwise, as did Penrose. At any rate, if you truly believe this and understand what you have just claimed, then you necessarily have just conceded that the Theorem does not say anything about the human mind and mechanism.

Nomos wrote:

How then would you explain how we know things without turning ourselves into non-halting Turing machines?

I understand knowledge in animals to be computational states of brain tissue enabling animals to effectively deal with their environments. Some scientists differ over what "computational" ought to mean when applied to the brain, but the idea that information processing is involved is solidly entrenched.

Nosoul wrote:

Faustus, has Pierce's conception of Abductive Reasoning ever seemed like much of a problematic for the conception of AI? I mean, isn't by definition AI supposed to only be able to reason by induction & deduction, if not by deduction entirely alone? What if there really is such a thing as Abduction in reasoning? What special problems would this pose for the development of AI?

I don't think there's any official "definition" of how A.I. is to proceed. What you just described sounds like the old-school approach. A neural net, however, would behave in a manner somewhat like what Pierce described via "abductive reasoning".

Nomos wrote:

Of course any conclusion reached by abduction is uncertain, as with all conclusions reached by inference from premisses, and it's hard to see how a computational machine would deal with uncertainty.

Neural nets built to recognize faces have to be trained and make mistakes. Since their judgments are determined by connection strengths, they come in degrees—degrees of uncertainty, one could say.

Nomos wrote:

A 'spandrel' is Dennett's metaphor for a evolved feature which is not an adaptation.

No, it's Stephen Gould's term. And if you think Dennett believes consciousness is a spandrel, you absolutely do not understand him.

By Faustus (Brian Peterson)

Muxol wrote:

Do you mean that proof theory and model theory are in no way related? It seems as if that is what you're implying.

I was merely denying that the theorem had significant implications for epistemology in general, as expressed by the term "knowing". Or at least, I've never once personally encountered a scholar who thought it did, which could very well be nothing more than my own personal ignorance. At any rate, I'm dubious to say the least that such implications could successfully be drawn out. The theorem's domain of applicability is too narrow and technical to have much use beyond mathematics and logic.

By Faustus (Brian Peterson)