Godel Discussion Continued On BeliefNet by Faustus (09/2005):

Nosoul wrote, *"But what about someone or something that makes mistakes, but nevertheless eventually at some point, if given enough time and giving enough effort, actually solves the algorithm?"*

This and some of your other points have prompted me to write further about the problems that come up with the argument of Penrose. I've usually deleted these points in the editing stages of previous essays I've written, but here goes. (For the record, my usual style of writing posts for Beliefnet is to write them as Word documents, save them, review them a few times and make changes—sometimes taking weeks--cut and paste into Notepad to get rid of formatting codes, then cut and paste into Beliefnet. Can't recommend this method enough; helps cut down on the spelling and grammar errors, and vastly improves the writing style, if not the strength of the argument. But I digress.)

Since Penrose's argument is based on the logical calculus of Godel, the propositions about the world he puts into the Godel grinder have to be clear, precise, not in need of interpretation, and unambiguously true or false. These are absolute requirements. Penrose fails completely unless he can prove that human beings are capable of a feat known to be impossible for any algorithm—the ability to prove true their own Godel-sentences.

Trouble begins the moment you stop thinking about this abstractly and figure out what it would take for a person to really perform such a task. The first problem is trying to figure out what a person's Godel sentence would be. It would literally have to be the Godel sentence of the person's "program", her algorithm. (At first glance, this appears to give the game away—for here we are, assuming people are algorithms in order to show they aren't. But what Penrose proposes instead is that due to unknown quantum mechanical principles that aren't algorithmic, the algorithm that is you gets a magic boost, giving you a kind of superpower to transcend the limitations placed by Godel on your algorithm. That magic quantum element is responsible for some kind of "essential" property of consciousness not covered by the parts of you that are subject to a computational approach.)

So, what would your Godel sentence look like, and what are you supposed to be able to do with it? It would literally be a line of programming logic with a Godel number tacked on the beginning. I'm wishy washy on whether your task is to somehow prove that it's true, or prove that it is consistent. Maybe it's neither. Doesn't much matter relative to the point I'm trying to bring out, which starts to become clear when you try to figure out what's in the code, and what it would look like. First, there is ambiguity about what should be put into this line of code. Presumably, it would have to describe the entire physical state of your brain using laws of physics suitably described in computer language. Should it include the rest of your body, as well, or can everything out in the universe be bracketed and count only as input to the brain? This is an extremely important point, if only for the possibility—also due to the paradoxes of quantum physics—that the laws of physics included necessarily in your Godel sentence might require that the state of the nearby universe, or perhaps the entire universe, be part of your algorithm! That's just one reason why this question is important; I'll get to another one in a bit.

But since we have no answer to that question as yet, let's provisionally assume that all we need to include in your Godel sentence is the complete description of the informational processing ability of your brain, with the rest of your body, and the rest of the universe itself, appearing as inputs. This would require a long line of code able to instantiate all the possible connections between neurons in your brain (said number is greater than the number of atoms in the universe), as well as detailing every bit of behavior in every neuron in your brain. It furthermore would have to instantiate every single variable and effect from the aspects of the brain ruled by chaos, where microscopic differences so small as to be nearly invisible can have large effects at the macro level. So you can see that we are dealing with a line of programming code so long that it would, with moral certitude, be longer than the *combined* lines of code of any and all programs currently existing, be they weather simulators, videogames, or operating systems. We are already at a situation where individual human beings are incapable of reading through and understanding existing programs without the aid of a computer. I submit that simply reading through your own Godel sentence once, let alone analyzing

it, would likely take more time than the heat death of the universe, let along the lifetime of one person. So on strictly pragmatic grounds, this is an impossible task.

But wait—when couching arguments about logic and pure math, one is always permitted to imagine ideal circumstances, right? Now we return to the question of what should or shouldn't count as your Godel sentence, and which factors get plugged in as the rest of the universe providing inputs. A normal person eats, sleeps, and might be affected by the phases of the moon. A normal person has memory limitations, attention limitations, and her body and brain gradually get old, break down, and ultimately die. These limitations, paired with the length of the person's Godel sentence and the enormous task of processing it, clearly make it impossible for an ordinary person situated in the real world to solve his or her own Godel sentence.

But these limitations shouldn't get in the way, one might want to say. All we want to know is whether a person could, "if given enough time and given enough effort," process her own Godel sentence. So, if we restrict ourselves to toy universes where people never die, never have to sleep, never get fatigued, aren't hampered by the limitations of memory, and have infinite amounts of time, perhaps then we would have a universe where the task could be accomplished.

Perhaps, but now we have two significant problems for Penrose. The first point was always my original one, that there is no way even in principle to prove that human beings can pull this stuff off, only Penrose's hearty enthusiasm that it can. You can't overthrow the entire modern regime of consciousness studies with just that in hand. A related difficulty would be to ask if all human beings must be included in the set of potential Godel sentence solvers. Is it acceptable to rule out babies, young children, the insane, the mentally retarded, or the illiterate? Presumably all on this list are conscious in some sense, yet cannot pull the sword out of the stone. Then what gets to count as "putting enough effort" into the problem?

Second, it's quite reasonable to claim that by giving a person infinite time and eliminating mental attributes like attentional and memory limits, one has changed the actual algorithm of the person's mind, producing essentially a genius doppelganger extremely well suited for answering the Godel challenge. If that is the case, then the corresponding Godel sentence has to be amended to include each and every one of those changes that don't come in as inputs from the environment, or otherwise it is not the actual Godel sentence of the subject. But if this is done, Penrose's argument fails on the grounds that it could only apply to imaginary people possessing imaginary mental attributes, and not to people as they actually have been built by evolution, who have always been the one and only subjects of the consciousness studies he seeks to criticize. It's all fine and nice to construct a toy universe when arguing about pure math or pure physics, because a universe suitably constructed for such a task will not "leave out" anything important. In the case of human minds, however, you have to retain everything that is important about them in any and all thought experiments.

So, a proponent of Penrose has a formidable task. First, at least some empirical evidence or suggestion must be offered to suggest that a suitably motivated human can actually do what no algorithm can. Second, a description and defense of what humans "matter" and which do not must be mounted. The third option for the Penrose enthusiast is to carefully define which parts of a person's body count essentially as running his or her algorithm, and which count merely as inputs, and to do so in a way that bypasses the charge that something has been left out, or grouped on the wrong side of the algorithm/input equation.

And once this has been done, there are still a host of other problems. For instance, Godel's challenge only applies to algorithms that are internally consistent. So the next task would be to construct a defense of the rather odd notion that human beings are never inconstant, or that some human beings are never inconsistent, and that only these beings matter for purposes of this argument. I don't imagine anyone is seriously going to want to bite that one!

By Faustus (Brian Peterson)