Gdels Theorem

An incomplete guide to its use and abuse

T.Franzen

April 6, 2010

Gdel's theorem has enjoyed an unparalleled attention outside the narrow logico-mathematical world. Much of that is unwarranted. It has really no applications, except metaphorical ones, beyond the logic of formal systems. Not even in mathematics. Or maybe especially not in mathematics, which can go on as if it had never happened.

Gdel's theorem is about formal systems. Formal systems consist in strings which are manipulated according to formal rules, which by themselves can be formalized. Formal strings can be codified by integers, if this is done in a computable way, the codification itself can be codified by a number. Such codifications are referred to as Gdel numberings, and although technical and usually usurping an inordinate amount of space in any presentation of Gdels proof, they are conceptually trivial. What is subtle is the ability it allows of mapping meta-statements of a system into the system itself, allowing it so to speak 'to speak about itself'. Exactly how complete this mapping of meaning into formal statements is, constitute the object of inquiry of Gdel. The result is, hardly surprisingly, that it is, if sufficiently powerful, not quite complete. Why this should be surprising is in fact a surprise, something that might comes as a surprise to most spectators. There are a lot of subtleties going on, and the subtleties mostly concern some technical points. But of course the devil is always in the details, in particular the crucial technical details.

First there are axiomatizations, such as that of Euclidean geometry and the real numbers that do not suffer from incompleteness. The reason is that they truly only involve finite number of axioms, while the celebrated Peano Axioms (PA for short among the aficionados) involve an infinite axiom scheme, namely by the axiom of induction. Already here we may see the germ of the incompleteness theorem. It is no problem for us humans to conceive of the infinity of all those axioms in one go, but it is not easy, in fact impossible, to formalize this axiom scheme into a finite set of axioms. So already here we may ask ourselves, why are not humans superior to machines, machines that can be thought of as formal systems? Why do we have to go through the trouble of Gdels theorem, as does Penrose, to voice the obvious? Our ability to conceive of infinity is something we do not allow formal systems to do (or formal systems do not allow themselves to do so?). We do stack our cards in our favour. Then of course it is a metaphysical question in what sense we really fathom infinity? It is about thought feeding on itself, and even if thoughts are only about finite things, the fact that it is about any finite thing, makes us see the totality, which goes beyond any of its instances. The set of all finite sets is not a finite set. Still to all intents and purposes this constitute in our minds sound reasoning, be it by intuition, tradition or our belief in a Platonic reality. As a further illustration of this principle, if we disregard all but a finite number of the axioms of PA, the system is consistent. Any contradiction only involves a finite number of axioms, thus no contradictions can occur, as any 'finite' approximation is already consistent. From this we conclude that PA is consistent, although we cannot formalize this insight within PA.

The gist of Gdels proof is to exhibit a statement 'G' that says that it is not provable within the system. If this is true, the system is in particular consistent, because in an inconsistent system you can prove anything (incidentally also the consistency of the system). The point is that this meta-statement about the system can be mapped into the system itself. Thus if the system is assumed consistent, we have a true statement which cannot be proved within the system, because if it was it would contradict itself, and that cannot happen in consistent systems. In short by some mental acrobatics we have managed to put the Liars paradox to do some work. Now this is the first incompleteness theorem, from it, or rather its methods of proof, follows the second, which pinpoints that the consistency of a theory is something the theory by itself cannot prove[1]. Incidentally the proof of the second incompleteness theorem was only sketched, that being sufficient for the logical community[2] .

A consistent theory is overrated. Obviously we may use false axioms and exhibit a system that is internally consistent but only says falsehoods about the world. PA is consistent, to assume anything else would be just pedantic nit-picking according to the author, a symptom of a malaise brought about by over-exposure to logic. But we cannot formally prove this fact within the capabilities of PA. Thus to extend PA by adding the axiom that PA is not consistent, provides a consistent logical system, which obviously states false things about the world. (It says (or 'says'?) that part of the system is inconsistent, while the system as a whole is consistent!). The reader is obviously somewhat confused by this, what does it mean? How do you prove inconsistency? You exhibit a proof of A and $not - A$. Does that not mean that if a system is inconsistent, that you can prove it is inconsistent, by exhibiting such a compromising chain of deduction? Of course if a system is inconsistent it can prove anything, including its own inconsistency. What is a proof by the way? Is it enough that there is a proof (like the sound of a falling tree in a forest empty of ears?), or do you have to exhibit the proof and say here is a proof, in order to prove something? This act of showing can of course not be formalized. But if PA is consistent, there is no such compromising string, but what happens if you add it? What does that do? We cannot prove the negation of that string, so the new bigger system does not lead to a derivation of that string and its negation. But does that string not imply negations, or is that string not a real proof, but just a 'picture' of a proof which hence is ineffective? It is on this level of subtlety you need to ponder. It is a matter of technical subtlety not philosophical.

What is so disturbing about a system not being able to prove its consistency? If we doubt the soundness of a system, is that doubt being mollified by the fact that the system in principle says 'I am fine, trust me, I have just proved that I am fine'? After all inconsistent systems always prove their consistency. It ought to be a warning sign. It could

[1] According to legend, von Neumann, was the only one in the audience of the meeting who really understood Gdel, and who more or less immediately saw the implication, which was also apparent to Gdel himself.

[2] later a complete proof was written down a few years later in Hilbert-Bernays, die Grundlagen der Mathematik.

only be a matter of reassurance provided we were already convinced of the consistency of the system.

The quest for absolute certainty is a metaphysical ambition. As such it is doomed to be frustrated. Already the Greeks knew that in reasoning you had to start from something, to make a leap of faith, and in fact the faith in logical reasoning itself is also in the nature of a leap, except of course we have nothing to leap from. We always find ourselves in 'media res' and that we have to accept and go on from there. Why should we be so afraid of the void that preceded us? Why should we abhor infinite regress, when we are perfectly happy with infinite deductions in the other direction. Implicitly there is a pragmatic attitude at work. Keep on working and worry about the problems when they arise.

The author then makes the point that we need not be hung up on the self-referential idea in the proof, this can be circumvented, and he brings up the arguments using algorithms for the solution of Diophantine equations. But ultimately this of course is related to the spectacle of computably enumerable sets whose complements are not computably enumerable, a proof by Turing that hinges on the diagonal trick. Computable enumerable set is also a key idea in the way we think of formal system. We can computably enumerate all the proofs, and we can also computably decide whether a purported proof is a proof, but according to Tarski we cannot computably enumerate the true statements. Neither can we computably enumerate the unprovable statements, because to do so is to produce all the provable statements and reassure ourselves that it does not appear, but this is an infinite process.

Thus there are rather simple ideas involved, and the kind of reasoning applied is mainstream mathematical. However, the object seems to be mathematics itself, and thus there is a potentially powerful loop. However, this loop does not seem to have proved anything exciting about mathematics.

At the end the author discusses some other approaches to the incompleteness theorem that gives, or at least seems to give, a little more meat. Kolmogorov complexity is an example. By a simple counting argument one easily persuades oneself that most strings cannot be generated by strings of shorter lengths. There is thus the notion of irreducible complexity. The explanation of the string is itself. It only is a string, a truly random string, with no hidden structure. But how do we give an example of an irreducible string? This is is easy, any string of length one (or zero for that matter) is of course irreducible and cannot be 'explained' in terms of shorter strings. The implicit challenge is of course to exhibit an infinite number of irreducible strings, or at least irreducible strings of non-trivial length. But how do we not know that we some day will not come up with an explanation of sorts for a given string? Could it be that irreducible strings can only be seen collectively, never as individuals? The very fact that you look at a string individually makes it reducible? Because after all how can we produce truly random strings? Are we not unable to see them, to construct them? Certainly not by some algorithm, although some algorithms produce strings that satisfy most test for randomness. Should we use methods transcending mathematics and logic, such as the radioactive decay of a piece of matter? Should we list all the strings up to a certain lengths and perform a kind of Erasthones sieve, by simply writing down all programs (up to a certain size) and submit it to all shorter strings and see what comes out? But that is always in relation to some kind of formalization. Is randomness, like beauty, in the eye of the beholder? The author does not delve into such matters, instead he discusses as examples of unprovable statements, assertions as to the complexity of given strings.

To return to Gdels incompleteness. We saw that the germ of the insight was to be found in our transcendent principles. We can see all the finite sets in one go, or at least think we do; while this is not possible using the limited formalism provided by say PA. But of course we can add our principle, which we consider very sound, but then the incompleteness theorem indicates new problems of formalizing our insight into consistency. What this leads to is the introduction of new 'infinities', not just the infinities beyond the countable, as discovered by Cantor, but infinities in the sense of 'inaccessible cardinals'. As those infinities start to mount, the sense of soundness begins slowly to ebb away, at least to the mathematician. Such exercises, forced on the logician, seem to look suspiciously close to the scholastic proofs of Gods existence to the mathematician. (It might be worthwhile to point out that in this 'game' the mathematicians are the only spectators of what is going on, playing the role of the general public. The general public as such has no clue at all.)

The human mind, as well the most powerful computer, has obvious limitations as far as patience, energy and storage. Those obvious limitations have of course obvious implications on the limits beyond which humans may never venture. So why do we need Gdels theorem to tell us that, what we have known all along? A single human can only think of a finite number of numbers. There has only been a finite number of humans, and most likely there will only be a finite number of humans in the future, but that is not essential to the argument to follow. What is the biggest number ever thought about by humans? Can such a number be written down? But by the very act of pinpointing it, what prevents you from pinpointing an even bigger one by pointing at its successor? Does it mean that although there might be a biggest number it can never be exhibited, because it would by that very process cease to be the biggest number? Humans can think of a lot of concrete numbers, but it is beyond their ability to think of them as a set, because by grasping the totality they go one beyond. This kind of self-referentiality appears rather familiar. By knowing something we change it. Is this not what lies behind the notorious difficulty in making the multifarious world, including our own thinking, formal? And even if you do it systematically, there will always be something missing, and you have to climb yet another level. If so, is not the philosophical impact of Gdel rather mundane?

April 7, 2010 **Ulf Persson:** *Prof.em, Chalmers U.of Tech., Göteborg Sweden* ulfp@chalmers.se