PAPER On Gödel's Theorems in relationship to Philosophy of Science

This essay deals with Gödel's Theorems in relationship to Philosophy of Science; firstly, in outlining Ludwig Wittgenstein's position on the limits of philosophical truth that we can derive from Gödel (and how this in turn impacts modern-philosophical conceptions of science), and secondly, the deeper uncertainty about consciousness that Gödel's theorems point to, most notably elucidated by Sir Roger Penrose.

https://www.youtube.com/watch?v=0nOtLj8UYCw

This issue was recently discussed in a conversation between Roger Penrose, Federico Faggin and Bernardo Kastrup. One interesting divergence is that Kastrup and Faggin do not regard the wave-function collapse of quantum mechanics to be physically real, whereas Penrose, who I agree with, maintains that it is. Where I believe Faggin and Kastrup are understandably getting lost in the woodworks, is in relating their own (clearly Continental, with all the talk of ontology, phenomenology etc) philosophical background to modern science. Faggin highlights his own theory of consciousness as starting with the *a priori* assumption of consciousness as primary and free will as indisputable; this alone creates a struggle for Faggin's theory, where we've taken a concept like 'free will', one that has no falsifiable elements, and tried to comport it with a modern scientific framework that is one hand building from formal results and on the other hand from falsifiable hypotheses backed up with experimental data. This tendency to try and conflate or intermix complex philosophical topics with scientific theories and mathematical results itself is essentially similar firstly to what Gödel and consequently Tarski's work reveals, and secondly to what Wittgenstein states on the topic. On the other hand, Roger Penrose, in this video and throughout his career (most notably with *Emperor's New Mind*) is stating something altogether different: That, from his view, based off of Gödel's work, consciousness cannot be a computable function.

So we have two issues here, firstly, how it is that broad philosophical frameworks (such as the ones Faggin and Kastrup allude to) relate to *any* interpretation of science, and secondly, how Penrose himself interprets the scientific implications of Gödel.

Wittgenstein himself can offer tremendous clarity here: "P is the truth, say nothing but P" -- what do scientific results show us? Results. What do those results mean? As soon as we ask that question, we're not just doing

science anymore, we're in a different domain altogether -- this precise point gets lost, so we try to pose philosophic positions as scientific fact and extrapolate from scientific fact information that fact itself does not contain.

Wittgenstein offers a nuanced, but heavily criticised, perspective of Gödel's Theorem that echoes the above, the key offending passage being:
"I imagine someone asking my advice; he says: “I have constructed a proposition (I will use ‘P’ to designate it) in Russell’s symbolism, and by means of certain definitions and transformations it can be so interpreted that it says: ‘P is not provable in Russell’s system’. Must I not say that this proposition on the one hand is true, and on the other hand is unprovable? For suppose it were false; then it is true that it is provable. And that surely cannot be! And if it is proved, then it is proved that it is not provable. Thus it can only be true, but unprovable.”
Just as we ask: “‘provable’ in what system?”, so we must also ask: “‘true’ in what system?” ‘True in Russell’s system’ means, as was said: proved in Russell’s system; and ‘false in Russell’s system’ means: the opposite has been proved in Russell’s system. —Now what does your “suppose it is false” mean? In the Russell sense it means ‘suppose the opposite is proved in Russell’s system’; if that is your assumption, you will now presumably give up the interpretation that it is unprovable. And by ‘this interpretation’ I understand the translation into this English sentence. —If you assume that the proposition is provable in Russell’s system, that means it is true in the Russell sense, and the interpretation “P is not provable” again has to be given up. If you assume that the proposition is true in the Russell sense, the same thing follows. Further: if the proposition is supposed to be false in some other than the Russell sense, then it does not contradict this for it to be proved in Russell’s system. (What is called “losing” in chess may constitute winning in another game.)" - Remarks on the Foundations of Mathematics

The key claim put against Wittgenstein, primarily off the basis of this passage (and some rather spirited comments Wittgenstein made on Gödel), is that Wittgenstein misunderstood Gödel's Theorems and extracted false philosophical propositions from it, which I find rather ironic because it appears to me that Wittgenstein's whole project is predicated on avoiding precisely that: Extracting false philosophy from science. A more detailed look at Wittgenstein's thoughts on mathematics, *Leaving Mathematics As It Is: Wittgenstein's Later Philosophy of Mathematics*, a PhD thesis by Ryan Dawson, presents this reading far more clearly than I can:
"... Wittgenstein can be read as not criticising Gödel's proof itself. Instead Wittgenstein will be read as trying to put pressure upon or block certain

misleading interpretations of the proof's significance and so doing so without himself advocating or presupposing a dogmatic thesis (and so his remarks are not motivated by e.g. a thesis concerning truth or proof"
Chapter 10.1
https://ueaeprints.uea.ac.uk/id/eprint/57209/1/Leaving_Mathematics_As_It_Is_Wittgenstein's_Philosophy_of_Mathematics__130615.pdf

Floyd and Putnam (2000) put an argument to this effect forward, a popular refutation is as follows from Lampert (2011):
"This reason is mistaken because of the following reasons:

1. *Gödel does not agree with the assumptions Wittgenstein starts his argumentation in the second paragraph: Whether P = ΠP and ¬P = Π¬P are valid is just what is in question and the philosophical upshot of Gödel's proof is to have proven that these assumptions are wrong. This, indeed, is in conflict with Wittgenstein's philosophy. Yet, Wittgenstein has to argue against the premises of Gödel's proof (especially DEF. which itself strengthen the interpretation assumption), if he wants to stick to these assumptions. One cannot pertain to argue against the incompleteness proof by presupposing the falsehood of its conclusion.*
2. *Gödel does not start his argument, by presuming ¬P and reducing it to absurdity: Instead, he only reduces Π¬P and ΠP to absurdity, thus putting forward the undecidability thesis ¬Π¬P & ¬ΠP. And this he does without assuming interpretation assumption. Only his move from undecidability thesis towards the incompleteness theorem presupposes the interpretation assumption without hereby using RAA.*

According to any given interpretation, Wittgenstein's notorious remark on Gödel cannot be appreciated as revealing a "remarkable insight" of "great philosophical interest", because either it is understood as simply affirming what Gödel said or as a misguided critique of Gödel's proof. Wittgenstein's argumentation is no challenge for the Gödelian, yet Gödel's argumentation is a challenge for the Wittgensteinian.'
https://wab.uib.no/agora/tools/alws/collection-6-issue-1-article-6.annotate

What I think is missing, is the context for a lot of what Wittgenstein said, *vis-a-vis Wittgenstein responding to Bertrand Russell*; consider what Russell himself said of Gödel's Theorems. Firstly a quote of Russell's from a 1963 letter to Leon Henkin, which I got from a Quora answer:
“It is fifty years since I worked seriously at mathematical logic and almost the only work that I have read since that date is Gödel’s. I realized, of course, that Gödel’s work is of fundamental importance, but I was puzzled by it. It made me glad that I was no longer working at mathematical logic.

If a given set of axioms leads to a contradiction, it is clear that at least one of the axioms must be false. Does this apply to school-boys' arithmetic, and, if so, can we believe anything that we were taught in youth? Are we to think that 2 + 2 is not 4, but 4.001? Obviously, this is not what is intended. ... You note that we [Russell and Whitehead] were indifferent to attempts to prove that our axioms could not lead to contradictions. In this Gödel showed that we had been mistaken. But I thought that it must be impossible to prove that any given set of axioms does not lead to a contradiction, and, for that reason, I had payed little attention to Hilbert's work. Moreover, with the exception of the axiom of reducibility which I always regarded as a makeshift, our other axioms all seemed to me luminously self-evident. I did not see how anybody could deny, for instance, that q implies p or q, or that p or q implies q or p. ... If you can spare the time, I should like to know, roughly, how, in your opinion, ordinary mathematics—or, indeed, any deductive system—is affected by Gödel's work"
https://qr.ae/p2KgMb

Now consider this quote by Bertrand Russell posthumously attributed to him in the Addendum section of the fourth edition of *The Philosophy of Bertrand Russell* (1971), which I got from this Stack Exchange thread: *"Not long after the appearance of Principia Mathematica, Gödel propounded a new difficulty. He proved that, in any systematic logical language, there are propositions which can be stated, but cannot be either proved or disproved. This has been taken by many (not, I think, by Gödel) as a fatal objection to mathematical logic in the form which I and others had given to it. I have never been able to adopt this view. It is maintained by those who hold this view that no systematic logical theory can be true of everything. Oddly enough, they never apply this opinion to elementary everyday arithmetic. Until they do so, I consider that they may be ignored. I had always supposed that there are propositions in mathematical logic which can be stated, but neither proved nor disproved. Two of these had a fairly prominent place in Principia Mathematica—namely, the axiom of choice and the axiom of infinity. To many mathematical logicians, however, the destructive influence of Gödel's work appears much greater than it does to me and has been thought to require a great restriction in the scope of mathematical logic. ... I adhere to the view that one should make the best set of axioms that one can think of and believe in it unless and until actual contradictions appear."*
https://philosophy.stackexchange.com/questions/3951/did-russell-understand-g%C3%B6dels-incompleteness-theorems

Now consider Gödel's theorems themselves, which I described in my article *Philosophical Implications of Gödel's Incompleteness Theorems*

and Tarski's Undefinability Theorems (2024), in hindsight the conclusions I derived from the theorems in that article are themselves postulates I'm not so sure of, but the relevant bit is:
"It is important to note that Gödel's Theorems simply state that any formal system can either be complete, or consistent, but that it cannot be both. With Gödel's statement, it was proved that any sufficiently complex and effectively axiomatised formal system will have statements that it cannot prove, otherwise it will become inconsistent. On its own, despite consistent misinterpretations, Gödel's Theorems don't necessarily mean anything, all it does is highlight what was at the time a novel feature of all formal axiomatic systems that fulfil certain criteria (being complex, consistent, and effectively axiomatised).
That's it.
A common misconception is to think of Godel's statement as 'true but unprovable', but a truth-value cannot be defined within the system, simply because an 'unprovable formula' can't be represented arithmetically, it is like trying to imagine 'nothing', that's the whole point. Formal systems such as Peano-arithmetic are 'first order languages' that are incapable of proving everything, so second-order languages need to be used to effectively evaluate them. In the context of a first-order formal language such as arithmetic, it makes no sense to describe what is, or isn't, true, as the formal language is only capable of describing the logical consequences of a given set of axioms."
https://www.academia.edu/123244228/PAPER_Philosophic_Implications_of_G%C3%B6dels_Incompleteness_Theorems

From my perspective, Wittgenstein's statements follow clearly from Gödel's theorems, Wittgenstein merely used a common-language expression of the same idea (leading itself to the interpretation that Wittgenstein misunderstood Gödel). Gödel's construction is equivalent to a mathematical expression of 'this statement is a lie' (hence, Liar Paradox), Gödel shows us that this statement (in Gödelian numbers) cannot be proven *within* that formal system; *therefore any* notion of the provability of the whole system comes from a different perspective, that is, from a higher-order logic. This is no different to what Nagel and Newman (1958) have said on Gödel's proof:
"This imposing result of Godel's analysis should not be misunderstood: it does not exclude a meta-mathematical proof of the consistency of arithmetic. What it excludes is a proof of consistency that can be mirrored by the formal deductions of arithmetic. Meta-mathematical proofs of the consistency of arithmetic have, in fact, been constructed, notably by Gerhard Gentzen, a member of the Hilbert school, in 1936, and by others since then. ... But these meta-mathematical proofs cannot be

represented within the arithmetical calculus; and, since they are not finitistic, they do not achieve the proclaimed objectives of Hilbert's original program. ... The possibility of constructing a finitistic absolute proof of consistency for arithmetic is not excluded by Gödel's results. Gödel showed that no such proof is possible that can be represented within arithmetic. His argument does not eliminate the possibility of strictly finitistic proofs that cannot be represented within arithmetic. But no one today appears to have a clear idea of what a finitistic proof would be like that is not capable of formulation within arithmetic."
https://en.wikipedia.org/wiki/Hilbert%27s_second_problem#Modern_viewpoints_on_the_status_of_the_problem

Therefore, Russell's understanding *is* flawed, "*are we to think that 2 + 2 is 4, but 4.001?*" -- Gödel's proof demonstrates that the question makes no sense, within the context of Peano arithmetic 2 + 2 is *always* 4 simply as consequence of the axioms, *any question* as to whether or not PA 'is true' or 'is untrue' itself *makes no sense* **within** PA. Further to this, Russell's statements that "our other axioms all seemed to me luminously self-evident" and "I should like to know, roughly, how in your opinion, ordinary mathematics -- or, indeed, any deductive system -- is affected by Gödels work" -- Gödel's work effectively demonstrates that, in whichever manner our axioms are 'luminously self-evident', that 'self-evidence' is *not* itself proved *by* our axioms, and secondly, that this effectively limits the *entire scope* of both ordinary mathematics as well as any deductive system.

So what is Wittgenstein really saying?
"P is not provable in Russell's system" an equivalent to Gödel's mathematical construction of 'this statement is not true'
"*proposition on the one hand is true, and on the other hand is unprovable?"* This seems to me, the primary confusion, and the accusation of Wittgenstein confusing Gödel's conception of 'provability' with 'truth', but what Wittgenstein is subtly pointing out *is* the error of confusing 'provability' with 'truth'
"*Just as we ask: "'provable' in what system?", so we must also ask: "'true' in what system?"* This is where Wittgenstein makes it clear; when we *are* equating provability with truth, if (and when) we ask 'is this statement provable?' We must also ask 'provable in what system?' The same goes for 'true', just as Gödel demonstrates the limits of provability in formal systems, Wittgenstein is highlighting the limits of 'truth' in discourse by showing us *the statement itself makes no sense*, just as it would make no sense to ask *is Gödel's statement true within PA*
"*'True in Russell's system' means, as was said: proved in Russell's system;"* Wittgenstein here equates 'proved in PA' *to* 'true in PA' to to

show us that we *can't* equate the two
"Now what does your "suppose it is false" mean? In the Russell sense it means 'suppose the opposite is proved in Russell's system'; if that is your assumption, you will now presumably give up the interpretation that it is unprovable. And by 'this interpretation' I understand the translation into this English sentence. —If you assume that the proposition is provable in Russell's system, that means it is true in the Russell sense, and the interpretation "P is not provable" again has to be given up. If you assume that the proposition is true in the Russell sense, the same thing follows"
If you equate 'true' and 'provable' in PA, then the statement 'suppose the opposite is proved in Russell's system' *translates to* 'suppose it is false', which as Wittgenstein has already demonstrated, *makes no sense* in the sense of leading to a contradiction ("*suppose it were false, then it is true it is provable*"), this also hold for the inverse, if you equate 'true' and 'provable' in PA, assuming 'P is not provable' to be true also leads to a contradiction, *hence* Wittgenstein is showing us *we cannot* equate 'provable in PA' *with* 'true in PA'
"Further: if the proposition is supposed to be false in some other than the Russell sense, then it does not contradict this for it to be proved in Russell's system." This I believe directly relates to the mathematical project that Gödel disproved, that is the *Principia Mathematica's* goal for a full axiomatisation of mathematics -- Wittgenstein is simply stating that to regard the proposition as true or false *in* any sense is 'in some other than the Russell sense', that is *not* within Russell's system (which we can equate to PA, or any other first-order system) -- this is precisely what I think Putnam (2000) was outlining as important, and is no different to Nagel and Newman (1958) statement "*these meta-mathematical proofs cannot be represented within the arithmetical calculus*"

So, from the Wikipedia description of the second Incompleteness Theorem, the numerical equivalent of the formula *Cons(F)* "there is no natural number that codes a derivation of '0=1' from the axioms of *F*'", under general assumptions, "will not be provable in F" -- this is the mathematical equivalent of Wittgenstein's philosophical argument that his statement P cannot be true in the Russell sense, that is, in Russell's system.
https://en.wikipedia.org/wiki/G%C3%B6del%27s_incompleteness_theorems#Second_incompleteness_theorem

Another, far more generalised way to state all of the above is exactly how Tarski did with his Undefinability Theorems. Tarski posits that we can have an 'interpreted first-order language of arithmetic' N in which a sentence of first-order arithmetic L can be represented as 'true' or 'false'. Thus a

statement in N is a statement on the truth of a statement in L. With Gödel numbers, statements in L can be encoded as natural numbers. Tarski then denotes T as the set of L-statements that are true in N, and denotes T* to be the set of L-statements that are true in N, *encoded as Gödel numbers*, only to then demonstrate that there is no formula of first-order arithmetic that can define T*, that is, no formula that can define every L-statement that is true in N.

The Wikipedia article quite accurately displays the consequences of Tarski's Theorem:
"Informally, the theorem says that the concept of truth of first-order arithmetic statements cannot be defined by a formula in first-order arithmetic. This implies a major limitation on the scope of "self-representation". It is possible to define a formula True(N) whose extension is T but only by drawing on a metalanguage whose expressive power goes beyond that of L. For example, a truth predicate for first-order arithmetic can be defined in second-order arithmetic. However, this formula would only be able to define a truth predicate for formulas in the original language. To define a truth predicate for the metalanguage would require a still higher metametalanguage, and so on"*
https://en.wikipedia.org/wiki/Tarski%27s_undefinability_theorem

Thus, rightfully, as Wittgenstein elucidated, it makes no sense to question if something is true or not *within* a first-order theory, for any conception of truth, it is defined *outside* of the system it is being applied to. *This* precisely is where I think we err when it comes to modern discourse on science, *which is that* we take formal results *and extrapolate* from them, results that *don't necessarily* follow but are instead dependent on an altogether different framework from the conception of the formal results. Wittgenstein's own statements seem to sharply align with the essence of Tarski's Undefinability theorems, and seems to be a stark delineation from the view Russell took of Gödel's theorems, hardly surprisingly given how often Wittgenstein was directly responding to Russell.

Thus I think it stands to reason that Wittgenstein correctly understood the philosophical implications of Gödel's theorem, and that what he outlined is fundamentally similar to how Tarski related Gödelian numbers to syntactic conceptions of truth. Where Wittgenstein has been criticised, as in the case of Lampert (2011), it is done in essence by pointing out that Wittgenstein's assumptions do not agree with Gödel's and that more so "Gödel's proof is to have proven that these assumptions are wrong" (Lampert, 2011) -- what this perspective misses is precisely that Wittgenstein is using a similar line of argumentation to show the same thing i.e. that the assumptions are wrong, that one *cannot* conflate

provability with truth. Thus, Wittgenstein has been accused of doing the very opposite of what he did, what he did is draw a delineation between truth and provability, whereas he has been accused of confusing the two.

Furthermore, following from Wittgenstein's conception, it doesn't make sense to question if the wave-collapse (or any other number of things, such as the Big Bang) are 'real' or not; the mathematics itself lends itself to the wave-collapse or the Big Bang, trying to argue that it is an 'epistemic phenomena' will get nowhere, rather, all we can say of the wave-function collapse *is* what the results say; any attempt to reason *beyond that* is outside the realm of science, so relating abstract conceptions of consciousness and free will to modern quantum mechanics is *altogether unnecessary*. Wherever a new scientific advancement occurs, it represents a change in our knowledge and understanding, the 'horizons of our world', as a whole; fundamentally, a discovery *like that* of Darwin's theory, has *irrevocable consequences* on how we understand and interact with anything, so therefore must *necessarily* reflect and represent itself in the philosophies we use to describe and understand the world, analogous to confusing truth with provability, *where there is confusion* is in *confusing* philosophy *with* science.

Where this all becomes intensely interesting is in relationship to Penrose's own theories about consciousness, which Penrose himself touches upon in the first video. Where Faggin is making a *philosophical point* that itself does not reason necessarily from the science, Penrose on the other hand is making an altogether different claim about the implications of Gödel's result for consciousness, stating for a number of years that consciousness is not the result of a computable process.

Roger Penrose's results actually follow *perfectly* from John Lucas' interpretations (and might I add, Smullyan's) interpretation of Gödel and Tarski, let's see what Lucas has to say in his 1961 text *The Gödelian Argument*:
"He was thus able to circumvent the ban on self-reference, and find an arithmetical formula which ascribed a certain arithmetical property to a certain number, which turned out to be the coded expression of that self-same formula's being unprovable from Peano's axioms for arithmetic, or Elementary Number Theory as it is called. In this way he was able to construct a formula which, in effect, says of itself that it is unprovable from Peano's axioms: but in that case it must be true, for if it were not, it would not be unprovable, and so would be both provable and false"
From this, almost instinctively the mathematicians will pose the same criticism of Lucas as they did to Wittgenstein, that Lucas is confusing

provability with truth, but Lucas' argument is far more sophisticated than that.

"I thought I could apply this to the mechanist hypothesis that the human mind was, or could at least be represented by a Turing machine. If that were so, I argued, it would be comparable to a formal system, and its output comparable to the theorems, that is to say the provable formulae, of a formal system. And since we evidently are able to do elementary arithmetic, the formal system must include Elementary Number Theory, in which case there would be a Godelian formula which could not be proved in the formal system, but which was none the less true, and could be seen to be true by a competent mathematician who understood Godel's proof"

This is where the seas part, so to speak. Firstly, Lucas is specifically outlining his use of Gödel's proof in relation to 'mechanist claims' (which I believe is the missing context to the Wittgenstein quote). Secondly, and far more interestingly, the result that Lucas is describing *actually* follows from Tarski's argument, not Gödel (though it should be stated that Gödel had already outlined a draft proof of Tarski's proof before Tarski did, this is indicative of only one thing, that Tarski's results immediately derive from Gödel's results as soon as you step outside pure mathematics): 'could be seen to be true by a competent mathematician' *is no different* that us demonstrating the consistency of PA within a metamathematical framework (such as the ones Nagel and Newman mention (1958) on the consistency of arithmetic), that is *we are already interpreting* first-order language within metalanguage frameworks, *therefore*, according to Lucas, *we cannot be* a formal system 'represented by a Turing machine' as the 'mechanist hypothesis' posits -- that is, by our concept of 'truth' we are already utilising a framework that is itself *not* a formal system, therefore reflecting that *we* are not formal systems.

"Hence no representation of his mind by a Turing machine could be correct, since for any such representation there would be a Godelian formula which the Turing machine could not prove, and so could not produce as true, but which the mathematician could both see, and show, to be true"

63 years later, this point seems *woefully underappreciated,* except for by Penrose, who was ruthlessly criticised for his statement to this effect.

"Godel's theorem is paradoxical, it purports to show that the Godelian sentence is unprovable but true. But if it shows that the Godelian sentence is true, surely it has proved it, so that it is provable after all. The paradox is resolved by distinguishing probability-in-the-formal-system from the informal probability given by Godel's reasoning"

Lucas is demonstrating his own understanding Gödel here, as well as outlining the key delineation that Wittgenstein also outlined; that is, that Gödels results only appear confusing with the conflation of provability and truth.

Now we get to the heart of Lucas' argument:
"Rather than ask high-flown questions about the mind we can ask the mechanist the single question whether or not the machine that is proposed as a representation of the mind would affirm the Godelian sentence of its system. If the mechanist says that his machine will affirm the Godelian sentence, the mind then will know that it is inconsistent and will affirm anything, quite unlike the mind which is characteristically ***selective*** *in its intellectual output.*
If the mechanist says that his machine will not affirm the Godelian sentence, the mind then will know since there was at least one sentence it could not prove in its system it must be consistent: and knowing that, the mind will know that the machine's Godelian sentence is true, and thus will differ from the machine in its intellectual output. If the mechanist does not know what answer the machine would give to the Godelian question, he has not done his home-work properly, and should go away and try to find out before expecting us to take him seriously.
In asking the mechanist rather than the machine, we are making use of the fact that the issue is one of principle, not of practice. The mechanist is not putting forward actual machines which actually represent some human being's intellectual output, but is claiming instead that there could in principle be such a machine. He is inviting us to make an intellectual leap, extrapolating from various scientific theories and skating over many difficulties. He is quite entitled to do this. But having done this he is not entitled to be coy about his in-principle machine's intellectual capabilities or to refuse to answer embarrassing questions. The thought-experiment, once undertaken, must be thought through. And when it is thought through it is impaled on the horns of a dilemma. Either the machine can prove in its system the Godelian sentence or it cannot: if it can, it is inconsistent, and not equivalent to a mind: if it cannot, it is consistent, and the mind can therefore assert the Godelian sentence to be true. Either way the machine is not equivalent to the mind, and the mechanist thesis fails."
This is the key point that stands out from Lucas' argument: If the mind is a machine, and fits the definition of Turing-complete, with 'Godelian sentence' representing the statement of the Liar Paradox as encoded as a Gödelian number, to ask 'the mind' (the proposed computer here) if the Godelian sentence is true, and to get an affirmative answer, would indicate that the mind is not consistent and can affirm anything, whereas to give a negative answer would indicate that the mind is consistent -- therefore the mind knows the sentence is true, *whereas* the formal system itself *cannot*! Now, if the mind is a computer, to ask the mind 'prove this statement is not true' only to get a negative answer 'this statement is not true' *itself indicates* a *knowing* of truth -- that is, the *results of* Gödel are also *understood* insofar as they are understood in relation to a higher-

order conception of 'truth', this *presupposes* a syntactic definition of truth *by which* consistency itself is understood; indeed, this is *precisely* what Tarski's expanded results show us.

To affirm the Gödelian sentence, is for the mind (or computer) to affirm "this statement is not true", thereby proving its own inconsistency (at which point you have to contend that the mind is inconsistent and can affirm *any* statement, by which logic the *statement of its inconsistency* has no meaning), on the other horn, the mind (or computer) *cannot* negate "this statement is not true" *without* reference to a syntactic definition of truth -- why? What formalisation of a Gödelian number allows us to encode its own lack of existence? None! Therefore, for the mind to negate "this statement is not true", it would require a foresight to what it does not possess -- that is, a higher-order conception of what it does not have. In a deeper sense, Gödel and Tarski's results are intelligible precisely because we understand formal results not just as logical consequences of axioms *but* also in reference to some syntactic conception of truth from what Tarski himself terms 'a metalanguage'. *That itself* reveals that the mind is not reducible to a computable function, *specifically because* the mind's output reveals knowledge that is not itself characteristic of a formal process; if it were the characteristic of a formal process, we could not ourselves have any concept of 'truth' to relate our formal statement to, being limited to *only* thinking in first-order formulae. Thus, Lucas' argument extends beyond the scope of Gödel's work, but points to a similar construal as both Tarski's and Wittgenstein's -- where Tarski indicates a metalanguage to interpret the truth of a first-order language, and where Wittgenstein indicates the difference between provability and truth, Lucas is indicating the same difference between machine logic and our ability to interpret it, and *from that* Lucas is rightfully outlining the necessary consequence that the mind itelf is not a machine (at once analogous to Tarski's delineation between first- and higher-order logic, and Wittgenstein's delineation between 'provable in Russell's system' and 'true in Russell's system').

Ergo, Lucas' conclusions (and by extension, Penrose's) follow just as naturally from Gödel's results, as Tarski's and Wittgenstein's.

Where I think Wittgenstein, Lucas and Penrose have all been misunderstood, is in referring to the implications of Gödel in a nonmathematical sense, whereas the criterion of Gödel's own proof only relates to a narrow concept of provability within mathematical systems that meet certain criteria, the second you invoke any discussion of 'truth' you are at once outside the boundaries set by Gödel: the arguments outlined and put forward by Wittgenstein, Lucas and Penrose themselves

relate Gödel's result to a broader conception of both formal systems *and* the concept of provability (insofar as it relates to truth) -- in that sense their arguments are characteristically similar variations of Tarski's undefinability theorems, which themselves *follow* from Gödel's results.

In that sense, criticisms of all three seem to follow the straight line of reasoning that we cannot derive such philosophical implications from formal results. That is *precisely* what all three *already* understand: Wittgenstein, in highlighting the difference between 'provable' and 'true' in Russell's system, and Penrose and Lucas, in relating Gödel (and Tarski's result) to an obvious realisation about consciousness: Specifically, that *given what we know about how formal systems work, as proved by our results from formal systems, we know that the mind has behaviour that is itself* ***not*** *computer-like, how we know this itself is an open question --* ***just as*** *it is an open question, how it is that we can understand Gödel's results in the first place*.

In that context, neither Lucas' nor Penrose's argument is a mystical attempt at avoiding the question of consciousness, and is much the converse, by pointing out specifically *what we don't know about consciousness*, and the need that outlines for a better scientific description of consciousness. In a variety of other ways, Penrose has provided arguments for the deep interrelationship between physics and a new understanding of consciousness, but to finish off, consider this: That the limitations of biology, neuroscience and psychology themselves must themselves, if Lucas and Penrose's interpretation of Gödel holds any weight, be limited by precisely our lack of clarity around *what* consciousness is.

Thus, as a loose characterisation, from Wittgenstein's argument we firstly have the viewpoint that it makes no sense to doubt the truth of Gödel's proofs and secondly from Lucas and Penrose we have the viewpoint that Gödel's proofs indicate that consciousness is non-computable, *itself* indicated by how we at once *can understand* the results of the limits of formal systems, thereby indicating *that we are not one*, which leads us to the need for a modern, scientific understanding of consciousness.