Deflationism and the Gödel Phenomena: Reply to Ketland

NEIL TENNANT

I am not a deflationist. I believe that truth and falsity are substantial. The truth of a proposition consists in its having a constructive proof, or truthmaker. The falsity of a proposition consists in its having a constructive disproof, or falsitymaker. Such proofs and disproofs will need to be given *modulo* acceptable premisses. The choice of these premisses will depend on the discourse in question.

In the case of an arithmetical proposition, a truthmaker would be an intuitionistic proof from acceptable first principles, or axioms; and a falsitymaker would be an intuitionistic disproof from the same. For many an arithmetical proposition φ, one is not at present in possession of either a proof or a disproof of φ—indeed, one is not possessed of any effective method that one knows will eventually produce either a proof or a disproof of φ.[1] That, says the intuitionist, is why one should refrain from asserting $\varphi \vee \neg\varphi$.

In my paper 'Deflationism and the Gödel Phenomena', I was not wearing the hat of an anti-realist committed to a substantial notion of truth. Rather, I was playing devil's advocate. The devil—an unwitting client, for whom I was appearing *pro bono*—was the deflationist. I was seeking to show that a certain argument appealing to the orthodox treatment of Gödel sentences in arithmetic did not count against deflationism about truth. Other writers, such as Shapiro and Ketland, had claimed that deflationism about truth could be discredited by consideration of the special status of independent Gödel sentences. Ketland, in particular, had written (1999, p. 88)

[1] This weak statement is all that one needs in this context. But of course we can make a stronger statement. We know (by Church's Thesis) that, for sufficiently powerful and consistent systems *S* of arithmetic, *there is no* effective method that will, for any given arithmetical sentence φ, eventually produce either a proof or a disproof of φ.

Mind, Vol. 114 . 453 . January 2005
doi:10.1093/mind/fzi089

> … our ability to recognize the truth of Gödel sentences involves a theory of truth (Tarski's) which *significantly transcends the deflationary theories.* (Emphasis in the original.)

I was concerned simply to point out that Shapiro and Ketland were perhaps entering an over-hasty judgement against the deflationist. I showed that there were deflationarily licit ways of arriving at the conviction that one should assert a Gödel sentence rather than deny it. Ketland's main charge is now that I am guilty of theft over honest toil, because I simply assume certain reflection principles, allegedly without any justification. So I need to say something in order to convince the reader that adopting reflection principles, without any supposedly justificatory detour through theorizing about truth, is not wholly unjustified and unmotivated.

The dispute turns on how we, as users of a current axiomatic system *S* of formal arithmetic (which we hope and trust is consistent, but which of course will then be incomplete) 'come to see the truth of' any independent Gödel sentence for *S*. One comes to see the truth of a mathematical claim by proving it mathematically, from acceptable principles, and not necessarily by theorizing about truth. Why, the deflationist will want to know, should matters be any different with a Gödel sentence? Given a Gödel sentence φ independent of *S*, we have, by definition (assuming *S* is consistent),

> There is no *S*-proof of φ

and (assuming *S* is 1-consistent)

> There is no *S*-disproof of φ.

We know, however, that *S* (like any other axiomatic system) cannot be the last word in the codification of our intuitions concerning licit methods for forming proofs and disproofs. So we seek a natural extension of *S* that will be stronger but still licit, and which will be able to settle the question whether φ.

It is well known that there is a proof, in primitive recursive arithmetic (PRA), of the Gödel sentence φ itself from Con_S. (Indeed, even weaker systems than PRA will suffice for such a proof.) So whether one has Con_S by outright axiomatic stipulation, or as the conclusion of some yet further argument, one will be able to append that proof of the Gödel sentence φ from Con_S, and a proof of φ would be completed.

The time-honoured mathematicians' way of extending any system *S* has been to adopt as a new axiom the consistency statement Con_S for *S*. That then secures the Gödel sentence φ as a new theorem.

But for the philosopher, this would then raise the justificatory question of how we know that the system *S* is indeed consistent. It would not be a satisfactory answer to this question to put one's head down and carry on proving *S*-theorems, even if '0 = 1' is not among them. Putting one's head in the arithmetical sands of *S* affords no clear view of the whole *S*-horizon, as it were. *One* way (not necessarily the only one) to keep one's head above the sands of *S*, and to answer the justificatory question, would be to postulate some truth-theoretic principles from which one could prove that all theorems of *S* are true. It should then be evident—except, perhaps, to the dialetheist—that *S* is consistent. From this it would follow that one could assert Con_S.

Never mind that one could press a further justificatory question about the status of these truth-theoretic principles themselves! The truth-theorist takes them for granted, and has his sought (truth-theoretic) extension of the system *S*. This is the high route preferred by Ketland. It involves truck with a substantial conception of truth, in which the deflationist might well be an unwilling passenger. And it involves the theorist in commitment to very powerful principles, which, when applied to *S*, yield an extension of *S* that is of very much greater consistency strength than the system $S + Con_S$ itself.[2] This is a consideration that Ketland overlooks.

My client the deflationist prefers the low route. He simply reflects on his current *S*-bound methods of proof, and likes what he sees. He feels confident about them. He can express this confidence by means of various new *mathematical* principles, of varying logical strengths, and all eschewing any use of a truth-predicate. Strongest among them is the following *soundness principle*:

$$\mathrm{Prov}_S(\ulcorner\varphi\urcorner) \rightarrow \varphi.$$

The deflationist might well wish to adopt all instances of this schema. After all, if he was willing to assert any sentence φ for which he had furnished an S-proof, why not then also be willing to assert any sentence φ for which he can furnish a proof to the effect that the sentence φ can be furnished with an S-proof? Note that the proof in question can, without loss of generality, be taken to be formed within a very weak fragment of arithmetic, namely the aforementioned system PRA of

[2] As Ketland points out in his own footnote 3, his truth-theoretic extension of PA is intertranslatable with the second-order subsystem ACA of second-order arithmetic. By a result of Michael Rathjen (personal communication) ACA can prove every theorem in each system in a transfinite sequence, of length $\varepsilon_{\varepsilon_0}$ of ever-stronger systems of the form PA + Con(PA), PA+ Con(PA) + Con(PA +Con(PA)), … (taking unions at limit ordinals). Yet all we need here is the first of these!

primitive recursive arithmetic. That is to say, any provable claim of the form $\mathrm{Prov}_S(\ulcorner\varphi\urcorner)$ can be proved in PRA.

The deflationist's soundness principle, however, is stronger than what is needed for a proof of the Gödel sentence for *S*. He can get by with something a lot weaker. Much more austere would be to adopt all instances of the following schematic principle of *uniform primitive recursive reflection*, in which, note, the conditional is dominant:

$$\forall n\, \mathrm{Prov}_S(\ulcorner\psi(\underline{n})\urcorner) \rightarrow \forall m\psi(m), \text{ where } \psi \text{ is primitive recursive.}$$

It turns out that adding to *S* the schema of uniform primitive recursive reflection for *S* produces exactly the system $(S + Con_S)$. This is the weakest extension of *S* affording a proof of the Gödel sentence for *S*.

There is no truck here with a substantial notion of truth. All one is doing here is applying the reflective thought involved in the soundness principle, but now within the scope of a universal numerical quantifier, and confined to complex predicates ψ within which there are no unbounded numerical quantifications. It is surely a very modest way of expressing one's confidence in the resources of the system S for providing proofs of one's assertions. There is only reflection on one's present axiomatic and deductive commitments, and an attempt to articulate the systematic upshot of these commitments, by framing the rather weak-looking reflection principle just given. The truth-predicate does not even feature in these reflections. Instead, one has found a non-truth-theoretic way of expressing one's commitment to stand by one's earlier methods for justifying one's assertions. No further justification is needed for the new commitment made by expressing one's earlier commitments. As soon as one appreciates the process of reflection, and how its outcome is expressed by the reflection principle, one already has an explanation of why someone who accepts S should also accept all instances of the reflection principle.

One accepts with equanimity that by formulating that commitment (to one's earlier methods) as a new principle, one has thereby taken on essentially new commitments. For that is the lesson of the Gödel phenomena. The more we try to express our confidence in our licit methods of proofs, the more we extend them. The very concept 'licit method of proof' is indefinitely extensible. We do not need a substantial concept of truth in order to establish this result; and the result applies to all methods of proof, including those that eschew explicit truck with truth.

Ketland writes

> … if the base theory is something like PA and the truth axioms are Tarski's truth axioms, then the global reflection statement becomes *provable* in the

> truth-theoretic extension. So, the truth-theoretic extension of PA is a non-conservative extension. Hence, the notion of truth must be substantial. Hence, deflationism about truth is mistaken.

The deflationist can take this in his stride. For all that Ketland is saying is the following:

> … if the base theory is something like PA and the truth axioms are Tarski's truth axioms, then the global reflection statement becomes *provable* in this very powerful, formal truth-theoretic extension. So, this very powerful, formal truth-theoretic extension of PA is a non-conservative extension. Hence, the notion of truth answering to this very powerful, formal truth-theoretic extension of PA must be substantial. Hence, deflationism about the notion of truth answering to this very powerful, formal truth-theoretic extension of PA is mistaken.

The deflationist would reply, of course,

> Quite so. But my deflationism concerns the ordinary notion of truth, not this overly rich notion that you think you are either entitled to or obliged to deploy, and to capture which you have formulated your very powerful, formal truth-theoretic extension of PA. If ever I do need to deploy the notion of truth, it will be a much more modest notion, answering only to the T-schema '$T(\ulcorner\varphi\urcorner)\leftrightarrow\varphi$'. But please note that I do not even deploy this weak notion when it comes to justifying my assertion of Gödel sentences for systems of arithmetic to whose axioms and rules I am committed. (Indeed, the T-schema would not suffice for that purpose!) My preferred method of justification proceeds without any mention of truth whatsoever.

Note, moreover, that the last-quoted claim of Ketland does not speak to my defence of deflationism against the earlier claim that Ketland had made and that I quoted above:

> … our ability to recognize the truth of Gödel sentences involves a theory of truth (Tarski's) which *significantly transcends the deflationary theories.*

It was to *this* last claim that my exploration of reflection principles, on behalf of the deflationist, was addressed. I showed that our ability to recognize that we should assert (rather than deny) Gödel sentences is intact even when we give up talking about truth altogether.

Ketland writes

> … Tennant's 'semantic argument' can be phrased in just one line:
>
> **Semantical argument:** If S is *sound*, then the Gödel sentence G is true.

> Tennant appears to be arguing that the deflationist 'has properly deflationary means for attaining the insight that' *G follows from the soundness of* S. I find this misinterpretation of the debate bizarre.

So do I. But whose misinterpretation is at issue? First, the 'semantic argument' is not mine; I presented a passage from Dummett's writing as an example of the semantic argument, and referred to other places in the literature where similar forms of argument have appeared. I was concerned to show, contrary to popular opinion, that the deflationist could give the same 'line' of argument without making any explicit use of a truth-predicate. As for Ketland's complaint that

> Tennant provides no justification for the crucial premiss, that the theory S is *sound*

it appears to escape Ketland that the semantical argument does not need the full force of the claim that the theory *S* is sound. Let me make explicit what ought to have been clear, in context, in my statement of the semantical argument which Ketland quoted. Here, in full, is the original passage from my paper, with clarificatory material now added in square brackets:

> *G* is a universally quantified sentence (as it happens, one of Goldbach type, that is, a universal quantification of a *primitive-recursive* predicate). Every numerical instance of that predicate is provable in the system *S*. (This claim requires a subargument exploiting Gödel-numbering and the representability in *S* of recursive properties.) Proof [of any such numerical instance] in *S* guarantees *truth*. Hence every numerical instance of *G* is *true*. So, since *G* is simply the universal quantification over those numerical instances, it too must be *true*.

Clearly the semantical argument needs, as its 'crucial premiss', no more than that all provable numerical instances of primitive-recursive predicates are true. This is the principle that I called (*pa*) in 2002, at p. 569. This principle is much weaker than the soundness of S, that is, than the claim that every theorem of S is true. Note, for example, that even the unsound (but presumably[3] consistent) system $\mathrm{PA}+\neg Con_{\mathrm{PA}}$ proves the same numerical instances of primitive-recursive predicates as does the sound system $\mathrm{PA}+Con_{\mathrm{PA}}$. In my faithful recapitulation, on behalf of the deflationist, of the deductive structure of the semantical argument, the reflection principle (*pa*) is not itself directly applied. Rather, the would-be application of (*pa*) becomes the following feature (i) of an application of uniform primitive recursive reflection. This application of uniform primitive recursive reflection deals at one blow with (i) the (*pa*)-

[3] I say 'presumably' here because the consistency of $\mathrm{PA}+\neg Con_{PA}$ follows from that of PA; and PA is, presumably, consistent.

style transition, for any numerical instance *n* of the primitive recursive predicate '$\neg Proof_S(\underline{n}, \ulcorner G \urcorner)$', from the statement of the provability of that instance to that instance itself; and (ii) the transition from the arbitrary instance '$\neg Proof_S(\underline{n}, \ulcorner G \urcorner)$' to the universally quantified statement '$\forall n \neg Proof_S(\underline{n}, \ulcorner G \urcorner)$', that is, to G itself.

By misrepresenting the force of the soundness claim adverted to in my statement of the semantical argument, Ketland is forcing into my unwilling hands a hammer to crack a walnut. He has not attended to a constant refrain in my paper: that the deflationist's reconstruction of the semantical argument must be faithful to its deductive structure—that it must preserve the 'line of reasoning' employed therein. So I am not impressed by his asseveration that Shapiro, Feferman and he do not *assume* that proof in *S* guarantees truth, but rather they *prove* it. So what? The philosophical point is that *they do not need to*, at least for the purpose of showing that one ought to assert the Gödel sentence for *S*, rather than deny it. (Here it should be borne in mind that my whole paper was focusing on that limited purpose.)

Indeed, it is at and on this very point that one can take the philosophical argument to the anti-deflationist, and inflict some real damage. Why should the anti-deflationist think that he has provided any sort of justification for the soundness claim when in doing so he appeals to theoretical principles that are much more powerful than the claim being justified? One is reminded of Tarski's reported quip, when asked whether Gentzen's proof of the consistency of arithmetic (using transfinite induction up to ε_0 on quantifier-free formulae) increased his confidence in the consistency of arithmetic. Tarski replied 'By about an epsilon.'

Ketland's comparison of the deflationist's position here with that of an instrumentalist in the case of empirical theorizing is not well taken. In the empirical case, one can give a compelling methodological justification for the need to state one's explanatory hypotheses at a level of greater generality and scope. This is because one is theorizing about the concrete and the contingent, and formulating hypotheses that are intended to be, at least in principle, *empirically falsifiable*. But in the mathematical case one is stating axioms (and proving theorems from them) about abstract domains, concerning which those theorems are *necessary*. In this context, justifications are *weaker* to the extent that the principles that they invoke are logico-mathematically *stronger*. Hence my unwillingness to use truth-theoretic hammers to crack mathematical walnuts.

Nor is Ketland right in claiming

> … *if* Tarski's theory of truth provides at least *one* way of 'recognizing the truth of Gödel sentences', then this fact alone contradicts deflationism (for a deflationary theory of truth should be conservative). The entirely different assumption, which I did not make, that this 'is the only way' is irrelevant.

This is deeply puzzling. It is obvious that I have always conceded the antecedent of the conditional: Tarski's theory of truth does indeed provide a way of recognizing the truth of Gödel sentences. My sole concern in my paper was to argue that this was not the only way of recognizing the truth of Gödel sentences. I sought (and found) a way for which the deflationist could opt, because it eschewed truth-talk altogether. I would not have succeeded in doing so had the Tarskian way been the only way—that is, if 'our ability to recognize the truth of Gödel sentences involves [Tarski's] theory of truth' (to quote Ketland yet again). It was crucially relevant to show that our ability in this regard need not be explained or analysed by appeal to a substantial concept of truth. Ketland's way of putting the matter, however, even when read carefully in context, makes the claim of exclusivity—that going via Tarski's theory of truth is the only way. Now, whether that claim of exclusivity was an assumption, or the conclusion of an argument, for Ketland, is neither here nor there; for Ketland made the claim. He now calls it an assumption, and says he did not make it. He also claims now that it is irrelevant. This I cannot fathom.

Ketland quotes Feferman in order to show how Feferman proved general reflection principles for *S* by using a suitably formalized truth theory for *S*. That only shows, however, that a way of proving reflection principles is to invoke some truth theory. It does not show that there is no other way to be justified in postulating those reflection principles—such as engaging in suitable intellectual reflection. Thanks to Feferman, one can regard the theorizing about truth as a ladder to be kicked away by the deflationist who has discovered (in the reflection principles themselves) other means of ascent.

Department of Philosophy
230 North Oval
The Ohio State University
Columbus, OH 43210
USA

NEIL TENNANT

