MICHAEL DETLEFSEN*

ON INTERPRETING GÖDEL'S SECOND THEOREM

1.

In this paper I critically evaluate the most widespread philosophical interpretations of Gödel's Second Incompleteness Theorem. My approach is to say what I think is wrong with these interpretations as they presently stand, and, where possible, to try to indicate what would have to be achieved were those interpretations to be revived, though revival is not, in my opinion, a reasonable hope.

Sections 2–7 discuss that cluster of interpretations that I choose to call the sceptical interpretations of Gödel's Second Theorem (hereafter G2). In Section 8 I consider that interpretation of G2 which attributes its significance to some alleged ill effects it has on Hilbert's Program. I shall argue there that G2 does not imply the failure of Hilbert's Program.

2. SCEPTICAL INTERPRETATIONS OF G2

There are many who have taken the position that G2 somehow shows that any consistency proof for a theory T of which it (G2) holds will have to make use of a premise-set that is more dubitable[1] than the premise-set of any proof constructible in T. Among those holding this position (or one similar to it) are E. W. Beth, Paul Cohen, A. Grzegorczyk, E. Nagel and J. R. Newman, and, most recently, M. D. Resnik. To illustrate this point I will cite passages from the writings of Beth, Cohen and Resnik, for in their writing we find unusually concise statements of the interpretation in question.[2]

In Beth and Cohen (respectively) we find the following assertions.

> ... according to this theorem [G2], the arguments needed in a consistency proof for a deductive theory are always in some respect less elementary than those admitted in the theory itself ...[3]

> ... [G2] implies that the consistency of a mathematical system cannot be proved except by methods more powerful than those of the system itself ...[4,5]

Journal of Philosophical Logic 8 (1979) 297–313. 0022–3611/79/0083–0297$01.70.
Copyright © 1979 *by D. Reidel Publishing Co., Dordrecht, Holland, and Boston, U.S.A.*

If taken literally, the claims of both Beth and Cohen are false. For surely G2 shows nothing about every 'deductive theory' or 'mathematical system', but, at most, shows something about those theories or systems to which it applies.[6] But if we allow Beth and Cohen this obvious restriction, their position is not patently false.

Neither, it may be thought, is their position patently epistemological. Why, it might be asked, should we take their remarks to be of an epistemological rather than a purely logical character?

My response is that it is more charitable to Beth and Cohen to take them as attempting to make an epistemological point than to take them as attempting to make a purely logical point. For if the phrases 'less elementary than' and 'more powerful than' are reasonably interpreted as expressing a logical relationship, then the claims of Beth and Cohen are palpably false, whereas this is not so if the phrases in question are taken to express an epistemic relationship.

On a reasonable logical reading, the remarks of Beth and Cohen would amount to saying that G2 shows that the premise-set of any consistency proof for T (where T is a theory to which G2 applies) is *logically* (deductively) more powerful than T itself, where T_1 is logically more powerful than T_2 if and only if T_1 is an extension[7] of T_2 but T_2 is not an extension of T_1. This reading would have Beth and Cohen asserting the obviously false claim that G2 shows that the premise-set P of any consistency proof for T is an extension of T, but T is not an extension of P.

My suggestion is, then, that we interpret Beth and Cohen as making an epistemological claim equivalent to that given in the opening paragraph of this section.

M. D. Resnik has recently proposed a view of G2 which bears certain affinities to the Beth–Cohen view. Yet despite the similarities, there is also an important difference between Resnik's position and that of Beth and Cohen. This difference consists in the fact that Resnik restricts the supposed epistemological impact of G2 to a smaller class of theories than it would appear to be restricted to by Beth and Cohen. The restriction is explicit in the following passage from Resnik.

> We know that a 'non-pathological' consistency proof for a system S will use methods which are not available in S. When S is as strong a system as we are willing to entertain seriously, then a consistency proof for it will yield no epistemological gain.[8]

Because of this notable difference between Beth–Cohen and Resnik, the views will be given different criticisms.

3. IS THE BETH–COHEN INTERPRETATION EPISTEMOLOGICALLY INTERESTING?

If the Beth–Cohen interpretation is to serve as a basis for attributing epistemological importance to G2, it would appear to stand in need of supplementation, for taken by itself, it would seem to be of little or no epistemological interest. We must try, then, to uncover certain additional premises needed to span the distance between the Beth–Cohen view and some interesting epistemological conclusion.

The conclusion toward which writers like Beth and Cohen[9] seem to be pressing is, as was mentioned in the opening section, what might properly be called a sceptical conclusion. For according to this conclusion there is an important restriction on how belief in the consistency of a theory might be justified; that is, such belief cannot be justified via proof of the usual formalizable variety. And, judging by our present lights, that amounts to saying that belief in the consistency of *T* cannot be justified by proof – period. It remains for us to try to fill the gap between the Beth–Cohen interpretation and this explicitly interesting conclusion in as plausible a way as seems possible. My suggestion is that, for want of an equally satisfactory alternative, we take the following pair of premises as furnishing the requisite supplementation.

SUP 1: If the premise-set of every consistency proof for *T* is more dubitable than the premise-set of any proof in *T*, then the premise-set of every consistency proof for *T* is more dubitable than the axiom-set for *T*.

SUP 2: If the premise-set of every consistency proof for *T* is more dubitable than the axiom-set of *T*, then the premise-set of every consistency proof for *T* is incapable of removing rational doubt or indecision from its conclusion.

If one supplements the Beth–Cohen view with the above pair of theses one gets to what looks like an interesting epistemological result. Without them, the epistemological importance of that view is not clear.

SUP 1 occupies a unique place in my criticism of Beth–Cohen. For either SUP 1 is an acceptable claim or it is not. If it is, then, as I shall shortly argue, it can be used to generate an attack on Beth–Cohen. If it is not, then Beth–Cohen is not worth attacking, since without the help of SUP 1, it is of little or no philosophical interest.

One could, I think, give quite a compelling critique of Beth–Cohen just by developing the above-mentioned dilemma. But my critique shall go further. I shall indeed argue for SUP 1. This I do with the thought that should the defender of Beth–Cohen fill the gap between it and some epistemologically interesting result (by means currently unforeseen) without appealing to SUP 1 (or some claim which entails it), then I would still have an argument against Beth–Cohen to fall back on.

4. AN ARGUMENT FOR SUP 1

The reader may already have observed that SUP 1 is trivially true for the class of finitely axiomatizable theories. If T is finitely axiomatizable, then the entire axiom-set of T will be the premise-set of various proofs in T. Thus, if the premise-set of any consistency proof for T is more dubitable than the premise-set of any proof in T, then it is, by that very fact, more dubitable than the axiom-set of T.

However, it is absolutely crucial to notice that the reasoning which has just been used to show the plausibility of SUP 1 for finitely axiomatizable theories cannot be used to show its plausibility in the case of non-finitely axiomatizable theories. The reason for this is plain: in non-finitely axiomatizable theories, the entire axiom-set is never the premise-set of a proof in the theory. Because of this, a non-finitely axiomatizable theory will have to be shown to possess a very special sort of epistemic organization if one is to be able to show that SUP 1 holds for it.

There is a special type of epistemic organization which T may possess and which enables us to demonstrate SUP 1 for T. This type of organization will be called 'epistemic compactness'.

We shall say that T is 'epistemically compact' when the dubitability of the whole theory is, in a sense, 'reflected' in a finite portion of the theory. More precisely, we shall say that T is epistemically compact when there is some finite subset T_f of the axioms of T that is as dubitable as the entire axiom-set of T.

The basis for my positive defense of the compactness thesis is the simple observation that often all one needs to know, and, in certain instances, all one can know, about a sentence A is information regarding its 'form' (e.g., its logical form, arithmetic form, set-theoretic form, etc.). Thus, in certain cases my only justification for believing a sentence A will be my belief that it has the (logical) form '$Bv\text{–}B$'. This will probably be true for cases where B is very long or complicated and for cases where B is some sentence undecided (and perhaps undecidable) by current knowledge. Similarly, I often commit myself to sentences whose content I've never examined and never will examine simply because they have a certain 'form'; for example, the 'form' prescribed by the axiom-schema of induction in Z, or the 'form' prescribed by the Aussonderungs principle of ZF. Indeed, tacit in the practice of using axiom-schemata to specify a theory is the assumption that the 'formal' information conveyed by the schema is both (i) all that we *need* to warrant acceptance of an instance of the schema and (ii) at least for certain instances, all the justification that we shall actually have for accepting those instances.

Now if we let '$In(AS)$' stand for an instance of axiom-schema AS of T that is justified and whose sole justification is the belief that it is of the 'form' prescribed by AS, then we get a reason for believing in the epistemic compactness of T. For under those circumstances, our confidence in '$In(AS)$' is only as high as our confidence in the claim that all instances of AS are true.[10] Thus, our confidence in $\{\{T_n\} \cup In(AS)\}$ will be no greater than our confidence in T entire. ($\{T_n\}$ is the set of axioms of T not given by AS).

The preceding argument gives us the means to defend SUP 1 for non-finitely axiomatizable theories that are compact.[11] For if T is given by the individual axioms $A_1, \ldots, A_n$ and the axiom-schemata $AS_1, \ldots, AS_k$, then the finite set T_f of axioms of T comprised of $A_1, \ldots, A_n$ and $In(AS_1), \ldots, In(AS_k)$ (where each $In(AS_i)$ is an instance of AS_i whose sole justification is the general claim that all instances of AS_i are true) is as dubitable as the axiom-set of T. Furthermore, T_f is the premise-set of some proof in T. So if the premise-set of every consistency proof for T is more dubitable than the premise-set of every proof in T, then the premise-set of every consistency proof for T is more dubitable than T_f. And if the premise-set of every consistency proof for T is more dubitable than T_f, then the premise-set of every consistency proof for T is more dubitable than the axiom-set for T. Thus, SUP 1 is true for finitely axiomatizable theories and also for non-finitely axiomatizable theories that are epistemically compact.

Now there is another type of epistemic organization which a non-finitely axiomatizable T may possess which does not entail T's compactness. I shall call this sort of organization 'epistemic paracompactness'.

Think of a non-finitely axiomatizable theory T which has a set $\{T_n\}$ of individual axioms plus an axiom-schema AS. And let $\{\{T_n\} \cup \{AS\}_j\}$ stand for the theory obtained by adding all of the instances up to and including the $j+1$th ($j \geqslant 0$) instance of AS to $\{T_n\}$. We may then define the sequence S to be the sequence where S_0 is the dubitability value of $\{T_n\}$ and where S_{m+1} is the dubitability value of $\{\{T_n\} \cup \{AS\}_m\}$. Then we shall say that T is epistemically paracompact if the sequence S converges at the dubitability value of T.

Now it would seem that S is a monotone increasing sequence since, surely, as we continue to pile instances of AS onto $\{T_n\}$ we keep getting sets of axioms of T that are at least as dubitable as the ones which preceded it in the 'piling on' process. Furthermore, S would seem to be bounded by the dubitability value of the entire axiom-set of T. And indeed the dubitability value of T's entire axiom-set might seem to form a least upper bound on S. But if this is true, then S converges at the dubitability value of T. And this being so, T is epistemically paracompact.

There are two points concerning paracompactness which I should now like to note. One is that the paracompactness of T does not sponsor a proof of SUP 1 for T, since for paracompact T, a set C may be more dubitable than any finite subset of axioms of T and still *not* be more dubitable than the entire axiom-set of T. Indeed, where T is paracompact but not compact, the entire axiom-set of T is itself such a C.

Secondly I should like to note that the paracompactness of T will turn out to be a strong enough condition on T to permit us to refute the Beth–Cohen thesis. This point will be developed in the next section.

Finally, before going on to a refutation of Beth–Cohen I should like to remind the reader that my overall strategy in this paper is to construct a dilemma for Beth–Cohen. I think that if Beth–Cohen is to be an interesting thesis, then one needs SUP 1. And if one needs SUP 1, then one needs epistemic compactness. If one finds insufficient reason to believe in compactness, then one should, to the extent of that insufficiency, doubt the significance of Beth–Cohen. By my argument involving compactness in the next section, all I am doing is trying to convince the reader that insofar as one has reason to believe in the compactness of T, one also has reason to reject Beth–Cohen as false. Of course, the reader may think that there is not good reason to believe in the compactness of T, and to that extent he may doubt my refutation of Beth–Cohen which appeals to compactness. But, since compactness would seem to be needed for Beth–Cohen to be significant, to the extent that the reader doubts my use of compactness to refute Beth–Cohen he should also doubt the significance of Beth–Cohen.

5. A CRITIQUE OF THE BETH–COHEN INTERPRETATION

My criticism of the Beth–Cohen interpretation will begin with a frontal attack, i.e., with an argument to the effect that it is literally false for a very wide range of mathematical theories. After finishing this phase of my criticism, I shall discuss various problems which will arise for one who might attempt to amend Beth–Cohen in such a way as to avoid the frontal attack.

But the frontal attack is of considerable importance because, if successful, it shows that the mere fact that C is a set of sentences which *logically implies* Con(T) cannot by any means be taken as evidence for the claim that C is more dubitable than the premise-set of any proof constructible in T. If there is a reason to believe that a given consistency proof for T will employ a premise-set P that is more dubitable than the premise-set of any proof constructible in T, that reason cannot consist in that mere fact that P logically implies Con(T).

The first argument that I would like to present is what (for reasons that will become apparent) I call the 'Reflexivity Argument'. This argument uses the reflexivity of Z, RA and ZF[12] and epistemic compactness to establish the falsity of Beth–Cohen for these theories. The argument is as follows.

(1) $Z(RA, ZF)$ is reflexive.

(2) By epistemic compactness, there is some finitely axiomatizable subtheory Z_f (RA_f, ZF_f) of Z (RA, ZF) such that the axiom-set of Z_f (RA_f, ZF_f) is as dubitable as the axiom-set of Z (RA, ZF).

(3) By (2) (and the way Z_f (RA_f, ZF_f) are constructed),[13] $\mathrm{Con}(Z_f)$ $(\mathrm{Con}(RA_f), \mathrm{Con}(ZF_f))$ is as dubitable as $\mathrm{Con}(Z)$ $(\mathrm{Con}(RA), \mathrm{Con}(ZF))$.

(4) By (1), $\vdash_Z \mathrm{Con}(Z_f)$ $(\vdash_{RA} \mathrm{Con}(RA_f), \vdash_{ZF} \mathrm{Con}(ZF_f))$.

(5) By (4), there is some finite set $\Delta_z(\Delta_{ra}, \Delta_{zf})$ of axioms of Z (RA, ZF) such that $\Delta_z \vdash \text{Con}(Z_f)$ $(\Delta_{ra} \vdash \text{Con}(RA_f)$, $\Delta_{zf} \vdash \text{Con}(ZF_f))$.

(6) By (5), $\text{Con}(Z_f)$ $(\text{Con}(RA_f), \text{Con}(ZF_f))$ is not more dubitable than $\Delta_z(\Delta_{ra}, \Delta_{zf})$.

(7) By (3), (6), $\text{Con}(Z)$ $(\text{Con}(RA), \text{Con}(ZF))$ is not more dubitable than $\Delta_z(\Delta_{ra}, \Delta_{zf})$.

(8) By (7), $\text{Con}(Z)$ $(\text{Con}(RA), \text{Con}(ZF))$ is not more dubitable than the premise-set of any proof in $Z(RA, ZF)$.

(9) By (8), Beth–Cohen is false for Z, RA, ZF.

One can also show the Beth–Cohen view to be literally false for a host of theories 'weaker' than Z. This class of theories is what I shall refer to as the finite Q extensions.[14] The argument is as follows.

(1) By the compactness of Z, there is some finite Q extension Q^* such that the axiom-set for Q^* is as dubitable as that for any finite Q extension.

(2) By the reflexivity of Z, the axiom-set of some finite Q extension Q^{**} (not identical to Q^*) logically implies $\text{Con}(Q^*)$.

(3) By (1), the axiom-set of Q^* is as dubitable as that of Q^{**}.

(4) By (2), (3), there is a proof of $\text{Con}(Q^*)$ whose premise-set is not more dubitable than the premise-set of every proof in Q^* (i.e., Beth–Cohen is false for Q^* and every finite Q extension that is an extension of Q^*).

I would now like to sketch an argument against Beth–Cohen *not* employing the notion of epistemic compactness. The argument that I have in mind appeals to only two features of T, namely (a) the paracompactness of T and the claim that (b) $\text{Con}(T)$ is at least slightly less dubitable than T. If (a) and (b) both hold, then there is some finite set of axioms T_f of T such that the dubitability of T_f comes closer to that of T than does the dubitability of $\text{Con}(T)$. That being so, one may take T_f and obtain the premise-set of a proof in T that is more dubitable than some set of sentences implying

Con(T). This, then, defeats Beth–Cohen even for theories that are *not* compact if they meet (a) and (b) above.

Now one may think that assuming T to meet (b) implies the falsity of Beth–Cohen for T. And this being so, it may be felt that the argument just given begs the question. But this is wrong. For (b) only implies the falsity of Beth–Cohen when (a) is present. In other words, one can only get from the assumption (which is (b)) that Con(T) is less dubitable than the entire axiom-set of T to the claim (which is the denial of Beth–Cohen for T) that some set of sentences implying Con(T) is *not* more dubitable than each premise-set of a proof in T, if one also has (a).

Two further points are worthy of notice. In the first place, it should be pointed out that, at least for a large class of philosophers of mathematics, namely, those who are platonists, (b) is an entirely reasonable assumption to make. For according to the platonist, there is more constraining truth in mathematics than mere consistency. Secondly, it should be noted that the argument from (a) and (b) is effective not only against a literal reading of Beth–Cohen, but also a reading of it which substitutes the phrase 'as dubitable as' for the somewhat stronger phrase 'more dubitable than'.

6. REVISING THE BETH–COHEN INTERPRETATION

A further look at the reflexivity argument suggests a strategy for revising the Beth–Cohen interpretation in such a way as to preserve its 'spirit' and at the same time free it from at least some of the difficulties discussed above. For the proof of Con(T) with which the reflexivity argument directly counters Beth–Cohen is the 'one liner' proof of Con(T). This proof is, of course, trivial in a certain epistemic sense, viz. that anyone having rational doubt or indecision concerning the truth of its conclusion will have as much rational doubt or indecision with respect to its premise-set.

Noting this, the advocate of the Beth–Cohen interpretation might attempt to restate his position along the following lines.

BCR: G2 shows that any 'non trivial' consistency proof for theory T will have to make use of a premise-set that is more dubitable than or as dubitable as the premise-set of any proof constructible in T.

Although BCR escapes the grasp of the Reflexivity Argument, it meets with other difficulties. To begin with, the argument concerning finite Q extensions given earlier, though not a clear-cut counter-example to BCR, is nonetheless a troublesome case. It raises a challenge to BCR, namely, to give reason for believing that the proofs of $\mathrm{Con}(Q^i)$ (where Q^i is either Q^* or a finite Q extension that is an extension of Q^*) are 'trivial'. If all such proofs are 'trivial', it is not in the least obvious that they are. As a result, until such a time as the advocate of BCR can show us that all of the proofs of $\mathrm{Con}(Q^i)$ are trivial, BCR will remain groundless for those cases of T comprised of Q^* and various of its extensions.

Another, and related problem confronting the advocate of BCR is to come up with support for BCR even in the case of Z, RA, ZF (and various extensions of each). For even in these cases BCR would appear to be groundless.

Grounds there would be if it could be shown that *every* set of sentences outside[15] of T which implies $\mathrm{Con}(T)$ is either as open to doubt as or more open to doubt than every set of sentences inside of T. But that thesis would seem to be falsified by $\{\mathrm{Con}(T)\}$ itself, as the Reflexivity Argument suggests.

And changing 'implies' to 'non-trivially implies' does not seem to help matters much. For there would seem to be no grounds of an *a priori* sort to suggest that just because a set of sentences lies on the exterior of T and 'non-trivially' implies $\mathrm{Con}(T)$ it will therefore be more dubitable than any finite set of sentences in the interior of T. Mathematical theories have not been consciously structured with such an end in mind.[16]

Nor would there appear to be any compelling empirical or inductive support for BCR. If we let A be 'G2 holds for T', B be 'all proofs of $\mathrm{Con}(T)$ are more dubitable than any proof in T' and C be 'all "non-trivial" proofs of $\mathrm{Con}(T)$ are more dubitable than or as dubitable as any proof in T', then it seems correct to say that $C(A, C) \not> C(A, B)$ (where $\ulcorner C(A, B) \urcorner$ is to be read 'the credibility of B given A').[17] This is due to the fact there are such results as Gentzen's consistency proof for Z and the Gödel–Gentzen proof of the consistency of Z using $\mathrm{Con}(H)$ and H and also due to the compound fact that BCR is challenged by the finite Q extension argument and that we see no reason to suppose that what happens with the finite Q extension won't also happen with the other theories for which G2 holds. In general, the problem for the advocates of BCR attempting an empirical defense is to show us that A is 'relevant' (epistemically) to C in a way that it is not 'relevant' to B. For the reasons just discussed, I think this would be quite difficult, if not impossible, to do for any appreciable range of theories for which G2 holds.

7. RESNIK'S INTERPRETATION

In Resnik's interpretation of G2 (cf. Section 2 of this paper for a statement of this view) the troublesome notion seems to be that of "a system as strong as we are willing to entertain seriously". Resnik makes the claim that for such a system S, a consistency proof (of the 'non-pathological' variety) will yield no epistemological gain. Resnik's claim is, I think, interesting because of the fact that it seeks to restrict the class of theories for which G2 has epistemological significance to those theories which are "as strong as we are willing to entertain seriously". To my knowledge, no comparable restriction is attempted by any other interpretation of G2.

That Resnik makes this studied restriction suggests that he thinks that there is something special about systems as strong as we are willing to entertain seriously that makes them, in a way not pertaining to other theories, epistemologically interesting targets for G2.

Let us call the theories to which Resnik restricts his claim R-theories. And let us see whether we can uncover any properties of R-theories that would serve to make them special.

One might want to characterize R-theories as those theories that are as dubitable as any theory which we are willing to entertain seriously. This is to give a decidedly epistemic reading to Resnik's phrase "as strong as". But so viewed, there would seem to be nothing separating R-theories from other theories in terms of the significance of G2. For, where T is as dubitable a theory as we are willing to entertain seriously, it would seem no more plausible to believe that there is no statement A outside of T that serves as a 'non-trivial' proof of Con(T), than it would to believe this where T is not such a theory.

What might seem to be a more refined view of what an R-theory is, is given by the following definition. T is an R-theory just in case (i) for every T' different from T if T' is (deductively) an extension of T, then T' is too dubitable to be entertained seriously and (ii) T is entertained seriously. The question then is whether for R-theories so characterized, it is plausible to believe that consistency proofs will yield no epistemological gain.

It would not seem possible to sustain a positive response to this question via some sort of *a priori* defense; i.e., by some defense attempting to use the mere notion of an R-theory plus an appeal to G2 to generate the claim that consistency proofs will yield no gain. For in order to construct such a defense, one needs *a priori* assurance that every set of sentences S outside of T which logically implies Con(T) does so 'trivially' (i.e., the initial dubitability of S is as great as or greater than that of Con(T)). But surely one has no such assurances *a priori*.

In fact, there is considerable reason to doubt whether, so characterized, there are any R-theories. One (but not, I think, the only) reason for doubting the existence of R-theories is the following. $T \cup$ Con(T) is an extension (deductively) of T, but it seems clear that the dubitability of $T \cup$ Con(T) is no greater than that of T. And this means that if T is a theory that we are willing to entertain seriously, then so also is $T \cup$ Con(T). If this is so, and it seems plausible enough, there could be no theory satisfying both clauses (i) and (ii) of the present definition of R-theories.

What Professor Resnik would seem to need is a theory T so constructed as to make every set of sentences on its exterior which implies Con(T) to be either (i) as dubitable as or more dubitable than any set of sentences in its interior or (ii) otherwise incapable of 'non-trivially' implying Con(T). But such a theory presents some special problems for Resnik. First, in what sense could such a theory be said to be a theory "as strong as we are willing to entertain seriously"? Secondly, what guarantee is there that such a theory would be of any importance to mathematics? Thirdly, and perhaps most importantly, what reason is there to think that such a theory would be recursively axiomatizable or otherwise capable of representing its proof theory and hence a theory for which G2 holds?[18]

Unless one can come up with the right answers to at least the latter two questions, it would seem that there is no hope of defensibly attributing epistemological significance to G2 on the grounds that there is some theory T whose exterior is (i) at each 'point' either more dubitable than or as dubitable as each 'point' of its interior or (ii) otherwise incapable of 'non-trivially' implying Con(T).[19]

And even if one could find the right answers to the latter two questions for a certain T, it is not clear that this would be of any aid to Resnik, for he would still be confronted with the task of showing for that T that, in some plausible sense, it was "as strong a system as we are willing to entertain seriously". As I doubt that this could be done, I also doubt that Resnik's restriction (to R-theories) is at all relevant to the quest for a plausible and epistemologically significant interpretation of G2.

8. G2 AND HILBERT'S PROGRAM

In this section I would like to sketch an argument against the claim that G2 implies the failure of Hilbert's Program for finding a finitistic consistency proof for the various theories of classical mathematics.[20] The central claim of the argument is that Con(T), the consistency formula shown to be unprovable by G2, does not really 'express' consistency in the sense of that term germane to an evaluation of Hilbert's Program.

In order for a consistency formula to 'express' consistency in the appropriate sense the quantifiers and operators in it must be construed finitistically, and *not* classically, since it is the finitistic consistency of a classical system that is at issue. But a finitistic interpretation of the universal quantifier would seem to differ drastically from a classical interpretation of it, as is clear from the following remark of Herbrand.

> ... when we say that an argument (or theorem) is true for all (these) x, we mean that, for each x taken by itself, it is possible to repeat the general argument in question, which should be considered to be merely the prototype of these particular arguments.[21]

And, again, he says that a proof of a universal claim is merely a description or manual of the operations which are to be executed in each particular case.[22] This view of the universal quantifier would seem to sponsor the following restricted w-rule: if I have an effective procedure P (i.e., a manual of operations P) for showing of each individual n that '$F(\bar{n})$' is finitistically provable, then '$(x)F(x)$' is also finitistically provable. Indeed in a 1930 paper,[23] Hilbert stated a rule something like this. And at that time it was apparent to finitists that the rule did not give one the power to go beyond the means of some methods that had already been accepted as finitistic.[24]

Now one would not, in general, want to add the above-mentioned w-rule to a scheme designed to serve as the finitistic proof theory of the classical theory T, since that rule does not constitute a truth of the finitistic proof theory of the classical T! Still, certain instances of the rule would seem to be called for; in particular the one producing Con(T) from its instances. This addition made, Con(Z) becomes provable in Z_{w^*} (= Z plus the above-mentioned instance of the restricted w-rule).

Of course, if one adds instances of the restricted *w*-rule to *T* in order to get an adequate context in which to do the finitistic proof theory of the classical *T*, then one will not be able to formulate the finitistic proof theory of *T* as a formal system, but I see nothing in Hilbert's Program which suggests that such formalizability is an essential or important feature of it. The essential thing is that *T* itself be formalizable, since if this is not the case, the consistency of *T* would not be a well defined finitistic problem.

G2, then, only seems to imply the failure of Hilbert's Program so long as one ignores the fact that the logic of the finitistic proof theory of the classical *T* and the logic of the classical *T* itself are two quite different logics! Once this is recognized, the fact that Con(*T*) is not provable in *T* should come as no particular shock to those espousing Hilbert's Program. If the logic of *T* is expanded in a way that produces a scheme whose logic is in agreement with the logic of the finitistic proof theory of the classical *T*, then in at least some instances (e.g., for the case where *T* is the system *Z*), Con(*T*) becomes provable. The basic flaw of those using G2 to thwart Hilbert's Program is that they fail to recognize that the logic of the arithmetized proof theory of *T* in G2 (since that arithmetized proof theory is itself embedded in *T*) is the logic of *T* itself, *not* the logic of the finitistic proof theory of *T* (which logic is *not* a subsystem of *T*'s logic)!

9. SUMMARY

In this paper I have considered various attempts to attribute significance to G2.[25] Two of these attempts (Beth–Cohen and the position maintaining that G2 shows the failure of Hilbert's Program), I have argued, are literally false. Two others (BCR and Resnik's Interpretation), I have argued, are groundless.

University of Minnesota–Duluth

NOTES

* I would like to thank Dale Gottlieb, Stephen Barker, Tim McCarthy, Philip Kitcher, Michael Resnik and Richmond Thomason for extensive and helpful discussion of this work.

1 My use of dubitability is mainly an intuitive one. That is, I make use of only those aspects of the notion which seem to lie at the core of our intuitions concerning

probability. I should, perhaps, also say that the notion of dubitability is taken here as applying to sentences and sets of sentences. When I speak of the dubitability of an interpreted sentence I intend to speak of the rational doubt and/or indecision concerning its truth. When I apply the notion to sets of interpreted sentences, I am intending to speak of the rational doubt and/or indecision concerning the joint truth of all sentences in the set.

I might also say that the notion of dubitability as I employ it could be derived from the notion of epistemic probability as characterized, say, in Skyrms' *Choice and Chance*. So, our judgments of the truth of sentences in mathematics may change with time, and the probability of sentences of mathematics is *not* taken to be (at least in all cases) either 0 or 1 depending upon whether the sentence is false or true. There is nothing inherently irrational (rational) *per se* about mathematical falsehood (truth) on my view.

[2] The appropriate sections in the writings of the other authors are [5], p. 576 and [7], p. 6 for Grzegorczyk and Nagel–Newman respectively.

Actually Grzegorczyk takes a somewhat stronger and Nagel–Newman a somewhat weaker position than the one stated above. I will not, in the course of this paper, explicitly discuss either of these variations. Suffice it to say that everything I say about the present interpretations clearly applies to the stronger claim of Grzegorczyk. For a thorough discussion of the apparently weaker claim of Nagel and Newman, I refer the reader to [4], pp. 71–78.

[3] [1], p. 74, brackets mine.

[4] [2], p. 3, brackets mine. Cohen makes a related remark in [3], cf. p. 13.

[5] In speaking of G2 as 'implying' such-and-such a conclusion C, Cohen can, I think, be taken as meaning that there is a set P of plausible statements such that P itself does not logically imply C, but $P \cup$ G2 does.

[6] For our purposes it will suffice to think of G2 as applying to the consistent, recursively axiomatizable extensions of the well-known system Q. However, the reader should be aware of the fact that G2 actually holds for a considerably wider class of theories. For a characterization of this broader class of theories see [9]. In the remainder of this paper when I speak of theories, I shall mean recursively axiomatizable theories, unless otherwise stated. I also take theories to be deductively closed sets of sentences.

[7] T_1 is an extension of T_2 iff both T_1 and T_2 are deductively closed sets of sentences and $T_2 \subseteq T_1$. If theories are not regarded as closed then T_1 can be said to be an extension of T_2 when the closure of T_2 is a subset of the closure of T_1.

[8] [8], p. 145.

[9] One finds a view like that of Beth–Cohen tempting Wang in his earlier writings, cf. [10], p. 27. However, Wang seems to have turned his back on this view in his more recent writing, cf. [11], pp. 42–43.

[10] An argument similar to this cannot be made when $In(AS)$ is not assumed to be justified. This is so because one circumstance that will lead to $In(AS)$'s not being justified is that the general claim is not justified (assuming, of course, that $In(AS)$ has no 'individualized' justification; i.e. no justification that applies to $In(AS)$ but not to the other instances of AS). And as there can be good reason for doubting the truth of the general claim (i.e., the claim that all instances of AS are true) that are not equally good reasons for doubting the truth of $In(AS)$, it cannot be expected that the dubitability of $In(AS)$ will match that of the general claim when $In(AS)$ is unjustified.

[11] The reader will recall that 'theories' for me means, unless otherwise stated, 'recursively axiomatizable theories'.

[12] A theory T is said to be reflexive just in case for every finitely axiomatizable subtheory T_f of T, $\vdash_T \text{Con}(T_f)$. Z (first-order number theory), RA (real arithmetic) and ZF (Zermelo–Fraenkel set theory) were first proved to be reflexive by A. Mostowski in [6].

[13] They are constructed by taking the finitely many individual axioms of $Z(RA, ZF)$ and adding to them for each axiom-schema in the theory, an instance of the schema whose sole justification is the claim that all of the instances of the schema are true, and then closing the set under deduction. That $\text{Con}(Z_f)$ will be as dubitable as $\text{Con}(Z)$ can be seen from the following argument.

Suppose that we divide the dubitability of a theory into two parts: (i) consistency worries and (ii) extra consistency worries, or 'factual' worries, as I shall call them. Now the factual worries concerning Z_f cannot exceed those of Z for every factual worry concerning Z_f is ipso facto one for Z. But, since the dubitability of Z_f equals that of Z (by epistemic compactness), and the 'factual' worries concerning Z_f cannot exceed those concerning Z, the consistency worries of Z_f must be equal to those of Z. In short, the consistency worries concerning Z_f cannot be greater than those for Z. And if they were less, the factual worries concerning Z_f would be greater than those concerning Z, which they cannot be. Thus, the consistency worries concerning Z_f equal those concerning Z. And what has been done for Z can obviously be repeated for RA and ZF. Thus we get (3).

[14] A finite Q extension will be any theory T meeting the following conditions.
(i) the axioms of Q are axioms of T
(ii) T has some (but not more than a finite number of) instances of the induction schema as axioms
(iii) T has only those axioms provided for by (i) and (ii).

[15] Since theories for me are deductively closed sets of sentences, a sentence being on the exterior of T (or outside of T) simply means that the sentence isn't a theorem of T. Similarly, but oppositely, for sentences in the interior of T.

[16] Indeed proofs such as Gentzen's of the consistency of Z which can be formalized in H with a constructivistic w-rule and Gentzen's and Gödel's of the relative consistency of Z to H (intuitionistic number theory) would seem to give at least some support to the view there are proofs of $\text{Con}(Z)$ on the exterior of Z which non-trivially imply $\text{Con}(Z)$ and which are not as open to rational doubt as every finite set of theorems of Z.

[17] Credibility is just to be treated as a cognate of dubitability. The more credible a sentence, the less dubitable it is.

[18] Rosser [9] showed that there are systems that can represent their proof theory even though these systems are not recursively axiomatizable using ordinary quantificational logic.

[19] A 'point' on the exterior of T is any set of sentences of the language of T that is not a subset of T. A 'point' in the interior of T is any set of sentences that is a subset of T.

[20] A full development of this argument, including historical remarks, is in a manuscript (available on request) which I am currently revising entitled 'Hilbert's Program and Gödel's Second Incompleteness Theorem'.

[21] In Goldfarb, *Jacques Herbrand: Logical Writings*, pp. 288–89, fn. 5.

[22] *Ibid.*, pp. 49–51.

[23] See Hilbert, 'Die Grundlegung der elementaren Zahlenlehre', reprinted in *Gesammelte Abhandlungen*, pp. 192–195.
[24] In Goldfarb, *Jacques Herbrand: Logical Writings*, p. 297.
[25] Conspicuously absent is any mention of how Feferman's work in his 1960 paper 'The Arithmetization of Metamathematics in a General Setting' might be called upon to reverse G2. I don't think it can, as I argued at the 1978 Western APA meeting in a paper entitled 'The Resolution of some Intensional Problems Concerning Gödel's Second Incompleteness Theorem'. That paper is currently being revised and expanded. A copy of the revised manuscript is available on request.

BIBLIOGRAPHY

[1] Beth, Evert W., *The Foundations of Mathematics*, 2nd ed. rev., North-Holland Publishing Co., Amsterdam, 1968.
[2] Cohen, Paul J., *Set Theory and the Continuum Hypothesis*, W. A. Benjamin, Inc., Reading, Mass., 1966.
[3] Cohen, Paul J., 'Comments on the Foundations of Set Theory'. In *Axiomatic Set Theory*, edited by Dana Scott, Proceedings of Symposia in Pure Mathematics, Vol. 13, pt. 1, The American Mathematical Society, New York, 1971.
[4] Detlefsen, Michael, 'The Importance of Gödel's Second Incompleteness Theorem for the Foundations of Mathematics'. Doctoral Thesis, The Johns Hopkins University, 1976.
[5] Grzegorczyk, Andrzej, *An Outline of Mathematical Logic*, D. Reidel Publishing Co., Boston, 1974.
[6] Mostowski, Andrzej, 'On Models of Axiomatic Systems'. *Fundamenta Mathematicae* 39 (1952), 133–156.
[7] Nagel, E. and Newman, J. R., *Gödel's Proof*, New York University Press, New York, 1958.
[8] Resnik, Michael D., 'The Philosophical Significance of Consistency Proofs'. *Journal of Philosophical Logic* 3 (1974), 133–147.
[9] Rosser, J. B., 'Gödel Theorems for Non-Constructive Logics', *Journal of Symbolic Logic* 2 (1937), 129–137.
[10] Wang, Hao, *Logic, Computers, and Sets*, Chelsea Publishing Co., New York, 1970.
[11] Wang, Hao, *From Mathematics to Philosophy*, Humanities Press, New York, 1974.