MICHAEL DETLEFSEN

ON AN ALLEGED REFUTATION OF HILBERT'S PROGRAM USING GÖDEL'S FIRST INCOMPLETENESS THEOREM

I. INTRODUCTION

For some time now, it has been a virtual constant of the literature on Hilbert's Program (HP) to maintain that Gödel's work demonstrates its untenability. The 'demonstration' typically given is one which proceeds from Gödel's Second Incompleteness Theorem (G2) and the claim that HP requires the sort of consistency proofs that it (i.e. G2) rules out. However, more recently (cf. Kreisel, 1976; Prawitz, 1981; Simpson, 1988; and Smorynski, 1977, 1985, 1988) it has become increasingly common to claim that Gödel's First Incompleteness Theorem (G1) affords a refutation of HP, and that this refutation is at least as good as (and perhaps even better than) that based on G2. Thus, one finds such claims as that "the First Incompleteness Theorem . . . effectively kills Hilbert's programme; the Second Incompleteness Theorem is merely a refinement" (cf. the introduction to Smorynski, 1988), and "it was the *First* and not the Second Incompleteness Theorem that killed Hilbert's Programme" (cf. 1985, p. 10).

Like the G2-based argument, the G1-based argument proceeds by producing a "requirement" for HP. This requirement is then shown to be ruled out by G1. In the case of the G1-based argument, however, the requirement is not one of consistency, but rather one stating that what HP identifies as 'ideal' mathematics must be a conservative extension of what it identifies as 'real' mathematics. G1 is then said to show that this conservation requirement cannot be met; the upshot being that G1 is sufficient, and hence G2 not necessary, for the destruction of HP.

We disagree with this argument. G1 does not 'kill' HP, and thus does not diminish the critical importance that the G2-based argument has for any attempt to establish a link between the defeasibility of HP and Gödel's work. In saying this, however, we are not intending to

Journal of Philosophical Logic **19**: 343–377, 1990.

suggest that the G2-based argument against HP is successful; indeed, we have argued at length elsewhere (cf. Detlefsen, 1986) to the contrary. Our aim is rather to affirm the pivotal status of the G2-based argument and thus to focus attention on it. We are thus making a proposal concerning what we take to be the proper focus of work aimed at determining the effect of Gödel's work on HP. The objectives of the present paper, then, are first to show that the G1-based argument is mistaken, and secondly to indicate what the main issues are that must be addressed by any rationally convincing evaluation of the bearing of Gödel's work on HP.

In connection with the first objective, we focus our attention on the role which the G1-based argument assigns to the conservation condition, and on the other alleged constraints on ideal theorizing which are then supposed to lead to its violation. We argue that both the conservation condition itself, as well as the other constraints on ideal theorizing which are taken to underlie its violation, are mistaken.

Regarding the second objective, our concern is mainly with the different conditions which G1 and G2 place on the (encoded) notions of proof, provability, etc. The conditions required by G2 are not satisfied by various theories or 'theory-like' arrangements of proofs and theorems which nonetheless do satisfy the conditions required by G1. Chief among these, so far as our interests are concerned, are various types of theories which incorporate consistency constraints into the very conditions on proof, provability, etc. We call these theories 'consistency-minded', and we maintain that some of them constitute plausible ways in which the Hilbertian might go about constructing his ideal theories.

There are two broad types of such systems; those introduced in Rosser (1936; and studied and refined in Kreisel and Takeuti, 1974; Guaspari and Solovay, 1979; Visser, 1989; and Arai, 1990), on the one hand, and those introduced by Feferman (1960) (and studied in Jeroslow, 1975; and Visser, 1989), on the other. Roughly speaking, the former (which we refer to as Rosser systems) begin with a class of would-be proofs and an ω-ordering defined on them, and require that, to be genuine, a would-be theorem must be determined neither to deny nor to be denied by any would-be theorem whose would-be proof precedes it in the given ordering of the would-be proofs.

[N.B. Since this requirement only involves comparison of the last line of the given would-be proof with that of its finitely many predecessors in the given ω-ordering, it is effectively executable.]

Similarly, the latter type of systems ('Feferman systems', in our terminology) make use of a class of would-be axioms and an ordering defined on them, and demands that for a would-be axiom to be genuine, it must be consistent with the whole class of would-be axioms which precede it in the given ordering. This in turn gives rise to a notion of proof which requires that, to be genuine, a would-be proof must be such that its 'largest' axiom (i.e. the one lying farthest out in the above-mentioned ordering) is consistent with the class of all "smaller" ones. [N.B. This means that, unlike Rosser systems, Feferman systems are not effective since, by Church's theorem, there is no effective way of carrying out the consistency check they require. However, in Jeroslow, 1975 and Visser, 1989, ways of obtaining effective systems based on Feferman-like ideas are suggested. Jeroslow calls these systems 'Experimental Logics' since their axioms are, in a sense, determined by a trial-and-error procedure.]

Because (at least some of) these consistency-minded theories seem to represent sensible strategies of theory construction for the Hilbertian, and because they at the same time violate certain of the conditions required for the proof of G2, they raise the question of whether these conditions are something to which the Hilbertian is committed by the nature of his enterprise. That these conditions hold for the usual sorts of systems but not for their consistency-minded counterparts should cause use to ask why the Hilbertian should be seen as committed to building his systems of beliefs in the usual static manner (where demonstrated consistency with previously accepted beliefs is not a condition on admitting a proposition as a new belief (i.e. theorem)) rather than in the more dynamic way suggested by the consistency-minded modes of construction. The answer, we believe, is that there is no reason; and this raises the question of whether G2 applies to Hilbert's Program *per se*, or only to those versions of it which needlessly restrict themselves to theory construction of the usual static variety.

We belive that this is an open question; and the primary objective of this paper is not to solve it, but rather to convince the reader of its seriousness, and to lay the philosophical groundwork necessary for a fruitful discussion of it. In this way, we hope to establish a general philosophical framework capable of serving as a guide to further work on the subject.

These, then, are the objectives of the paper. Its plan is as follows. In the next section we attend to preliminaries, setting forth a correct account of Hilbert's real/ideal distinction, which plays a central role in determining what sort of soundness condition it is appropriate for the Hilbertian to place on ideal mathematics. In section three, we give an analysis and critique of the G1-based refutation of HP. Finally, in section four, we attempt to lend some perspective to the refutation of the G1-based argument given in section three by showing what happens when attention is shifted back to the G2-based argument. Our belief is that the possibility of a consistency-minded conception of the Hilbertian's ideal theorizing may afford a way for the Hilbertian to carry out his program unimpeded by G2. At any rate, this is an alternative we think merits serious investigation, and whose philosophical groundwork we intend to prepare.

II. THE REAL/IDEAL DISTINCTION

In order to gain a proper understanding of the G1-based argument against HP, it is necessary to begin with a proper understanding of Hilbert's well-known (if not always well-understood) distinction between real and ideal mathematics. Hilbert himself seems to have intended this distinction to mirror the familiar distinction (of days gone by) between observation and theory in the natural sciences,[1] according to which observational evidence acts as a constraint on proper theorizing in this sense: no proper natural scientific theory can admit as consequences statements whose falsehood can be established by observational means. In this way, observation statements (i.e. statements whose truth is decidable by observational means) function to control theorizing in the natural sciences.

Construing the real/ideal distinction in parallel fashion, real proofs occupy the role of observational evidence, and ideal propositions the role of theoretical sentences so that the real proofs function to control ideal theorizing in the same way that observational evidence controls theorizing in natural science. Hence, just as no adequate natural scientific theory is permitted to have observationally falsifiable consequences, so too no proper ideal theory is permitted to have consequences whose falsehood can be established by real (i.e. finitary means).[2]

Hilbert seems to have viewed both natural scientific theory and ideal mathematics instrumentalistically. Thus, in drawing the analogy between ideal mathematics and theorizing in physics (1927, p. 475) he says that

> . . . a theory by its very nature is such that we do not need to fall back upon intuition or meaning in the midst of some argument. What the physicist demands precisely of a theory is that particular propositions be derived from laws of nature or hypotheses solely by inferences, hence on the basis of a pure formula game, without extraneous considerations being adduced.[3]

For Hilbert, then, ideal sentences do not have any genuine semantical or justificatory standing of their own. They do not express propositions, but are rather part of a formal calculary device (the system of ideal proofs of derivations) whose purpose is the timely and efficient derivation of sentences (viz. real sentences) that do have such semantic and epistemic standing.[4]

Smorynski (1988) is a reaction to this way of understanding the real/ideal distinction. His idea is to replace "the old dichotomy between finitary propositions and transfinite, or ideal formulae" with a new 'trichotomy' which distinguishes three types of propositions; namely, *real propositions*, *finitary general propositions*, and *ideal propositions* (cf. pp. 58 – 59). In this new trichotomy, the real propositions are identified with equations relating terms for primitive recursive functions and/or fixed numerals, and their finite propositional combinations. Their character is described as that of being "directly contentual assertions verifiable by direct computation". Their epistemological significance is that of "affording a control on the results of formal mathematical proofs". Thus, their epistemological role is like that of experimental or observational controls in the natural sciences, and they therefore play the same basic role played by the reals in the old dichotomous scheme.

The elements of the third category of the 'new' trichotomy – the so-called ideal propositions – are also conceived of in the usual way.

They are said not to be genuine propositions at all, but rather mere formulae and are said to bear the same relation to the reals as the theoretical elements of an instrumentalistically conceived natural science bear to the observation statements which pertain to it. Here, however, this relationship is apparently taken to be given not by the sort of *soundness condition* stated above (viz. that requiring that the theoretical elements not generate any observationally falsifiable consequent), but rather by a *conservation condition* requiring that any observational proposition derivable by theoretical means also be provable by observational means. At any rate, Smorynski maintains that in order for a given ideal theory **T** to be adequate, all real sentences provable in **T** must also be provable by real means (cf. p. 64). Of course, since the reals on Smorynski's scheme are all finitarily decidable propositions, there is no essential difference between this condition of real-conservation and the usual condition of real-soundness.

In its first and third categories, then, Smorynski's trichotomy follows fairly closely the usual view of the real/ideal distinction. The novelty comes with its second category, the so-called *finitary general propositions* (or 'fgp's', for short). These are generalizations such as 'for all numerals n, $n + 1 = 1 + n$'. According to Smorynski, the fgp's, like the real propositions, are supposed to be meaningful. Moreover, like the real propositions, they are said to be capable of finitary proof (albeit by what are called 'schematic methods'). What distinguishes them from real propositions, says Smorynski (cf. pp. 59–60), is that they are infinite conjunctions of real propositions, and, as such, function as the 'laws' of mathematical science. This is taken to be important because it is assumed that every genuine science must have laws (cf. Hilbert, 1925, p. 376, where it is said that the science of mathematics is neither exhausted by nor reducible to equalities and inequalities).

Fgp's thus seemingly have dual status; in their office as universal laws, they occupy the place of 'theoretical' elements, while in their role as objects of finitary proof, they function as 'controls' on ideal theorizing (and thus are to be included in the scope of the conservation condition as propositions which must be provable by finitary means whenever they are provable by ideal means).[5] We are thus left with the suggestion that the theoretical part of mathematics is a mixed bag; partly realist in character (owing to the genuineness of the finitary general propositions), and partly instrumentalist in character (owing to the merely 'computational' status of the so-called ideal propositions).

Many questions concerning this trichotomous conception of the real/ideal distinction are left unanswered by Smorynski (e.g. If the fgp's, like the reals, are subject to finitary proof, why is it only the latter and not the former which function as evidensory controls on ideal theorizing? If, on the other hand, the fgp's do not function as controls, then why should ideal theorizing be required to be conservative with respect to them? Moreover, are the fgp's propositions or propositional schemata? And how, exactly, do they compare to the ideal propositions?) Of more immediate concern than this, however, are the inaccuracies and confusions upon which it is built.

The chief culprit in all this is the introduction of the 'new' category of fgp's, which Smorynski sees as being based on two grounds. The first is the aforementioned idea that every genuine science must have laws (cf. pp. 59–60). Smorynski takes Hilbert to have been referring to a realm of 'laws' (viz. Smorynski's fgp's) when he spoke of that part of mathematics that cannot be reduced to numerical equations. In another place however, (cf. Hilbert, 1927, p. 471) Hilbert put his point more carefully. There he characterized mathematics not as the adjunction of an equational part and a set of contentual laws (Smorynski's fgp's) related to the equations as theory to data, but rather as the adjunction of real and ideal methods where the latter are related to the former as theory to data. Indeed, he says explicitly that what is essential for a genuine science of mathematics is not a realm of contentual laws (like Smorynski's fgp's), but rather the use of ideal methods (". . . scientific mathematics becomes possible only through the introduction of ideal propositions").

Thus it seems that what Hilbert had in mind when he spoke of a part of mathematics that is essential to its being a science and is not reducible to numerical equations is ideal mathematics, and not, as Smorynski suggests, a realm of contentual laws lying somewhere between ideal mathematics and elementary equalities and inequalities. Likewise, when he spoke of the equational part of mathematics as forming a control on the part of mathematics not reducible to it, this was just a loose, and somewhat sloppy, way of saying that the non-contentual part of mathematics (i.e. ideal mathematics) is to be constrained by the contentual part (including those reals that are not numerical equations as well as those that are!) in a way similar to that in which theory is to be constrained by observational data in the natural sciences.

This way of reading allows us to make sense of the previously quoted remark (Hilbert, 1927, p. 475) in which he compared theorizing in mathematics with theorizing in the natural sciences. There he stressed the idea that a theory "by its very nature" is supposed to eliminate the need "to fall back on intuition or meaning in the midst of an argument" and to allow argument to proceed "on the basis of a pure formula game". This is just another way of saying that what is essential to science is not contentual laws, but ideal propositions and derivations. Smorynski's fgp's, being contentual laws, could not function as part of a "pure formula game". Therefore, it is doubtful that they are what Hilbert had in mind when he spoke of an essential, non-equational component of the science of mathematics.

Smorynski's first ground for introducing the 'new' category of finitary general propositions is thus implausible. Still, it is better than his second one, which contends that the new category of fgp's is needed in order to make sense of Hilbert's views concerning the difference in meaningfulness between finitary general propositions, on the one hand, and existential generalizations, on the other. As is well-known, Hilbert emphasized the asymmetry of these two kinds of propositions, treating the former but not the latter as meaningful; an attitude which has often puzzled students of Hilbert (myself included). Smorynski claims that if one reads Hilbert as advocating a separate category (separate, that is, from the reals and ideals) for the fgp's, then one can make perfect sense of this otherwise puzzling view. The argument given for this claim is the following (pp. 59–60).

Real propositions do not exhaust the class of finitistic propositions. There are also what I shall call here the finitary general propositions — assertions of the form "for every numeral n, $n + 1 = 1 + n$", which Hilbert would have written in Leipzig as the free-variable formula,

$$x + 1 = 1 + x.$$

The general assertion "is from the finitist point of view *incapable of being negated*" because the negation ("for some numeral *n* . . .") "cannot be interpreted as a combination, formed by means of 'and', of infinitely many numerical equations". [It may seem odd that infinite conjunctions are finitistic assertions. However, i. one cannot seriously consider any science which does not propose universal laws, ii. there are finitary schematic methods of proof, iii. consistency, which must be proven finitistically, is such an universal assertion, and iv. it was his *existential* theorems that Hilbert had been criticized for.

Anyone familiar with the passage from which Smorynski's paraphrase is wrested, will immediately see that something has gone wrong. Hilbert did *not* view finitary generalizations as infinite conjunctions; indeed, he presented infinite conjunctions as quintessential cases of *non-finitary* propositions. He therefore would not have used non-equivalence to an infinite conjunction as a standard for establishing the non-finitary character of existential generalizations. What Smorynski has done, then, is to produce an interpretation of the 'paraphrased' remarks that is not only different from, but, indeed, the very antithesis of what Hilbert in fact said.

This mistaken interpretation of Hilbert is the result of a simple misreading of the relevant texts. To locate the mistake, we shall quote the passage from which Smorynski's paraphrase is taken, first the German original and then its standard English translation.

Wir stoßen also hier auf das Transfinite durch Zerlegung einer existentialen Aussage, die sich nicht als eine Oder-Verknupfung deuten läßt. Desgleichen kommen wir zu transfiniten Aussagen, wenn wir eine allgemeine, d.h. auf beliebige Zahlzeichen sich erstreckende Behauptung negieren. So ist z. B. die Aussage, daß, wenn $\mathfrak{a}$ ein Zahlzeichen ist, stets

$$\mathfrak{a} + 1 = 1 + \mathfrak{a}$$

sein muß, vom finiten Standpunkt *nicht negationsfähig*. Dies können wir uns klar machen, indem wir bedenken, daß diese Aussage nicht als eine Verbindung unendlich vieler Zahlengleichen durch 'und' gedeutet werden darf, sondern nur als ein hypothetisches Urteil, welches etwas behauptet für den Fall, daß ein Zahlzeichen vorliegt.

This is the German original from *Mathematische Annalen*, 95 (1926), p. 173. Now follows Bauer-Mengelberg's translation from van Heijenoort (1967), p. 378, with which we agree.

Thus we encounter the transfinite when from an existential proposition we extract a partial proposition that cannot be regarded as a disjunction. In like manner we come upon a transfinite proposition when we negate a universal assertion, that is, one that extends to arbitrary numerals. So, for example, the proposition that, if n is a numeral, we must always have

$$n + 1 = 1 + n$$

is from the finitist point of view *incapable of being negated*. This will become clear for us if we reflect upon the fact that [from this point of view] the proposition cannot be interpreted as a combination, formed by means of 'and', of infinitely many numeral equations, but only as a hypothetical judgement that comes to assert something when a numeral is given.

The crucial phrase is ". . . diese Aussage nicht als eine Verbindung unendlich vieler Zahlgleichungen durch 'und' gedeutet werden darf" (translated as "the proposition cannot be interpreted as a combination, formed by means of 'and', of infinitely many numerical equations . . .", by Bauer-Mengelberg, although the 'the' would probably better – if less mellifluously – be translated as 'this'). And the crucial question regarding this phrase is this: To which 'Aussage' (proposition) is Hilbert referring?

The answer, of course, is that he is referring to the proposition "for every numeral n, $n + 1 = 1 + n$", and what he is saying about it is that it cannot be interpreted as an infinite conjunction. This simple and obvious reading, however, is directly opposed to Smorynski's. He takes the 'diese Aussage' of the passage quoted to refer not to "for every numeral n, $n + 1 = 1 + n$", but rather to its denial! Thus, according to Smorynski, what Hilbert was saying when he wrote ". . . diese Aussage nicht als eine Verbindung unendlich vieler Zahlengleichungen durch 'und' gedeutet werden darf" is that the unbounded existential proposition "for some numeral n, $n + 1 \neq 1 + n$" (i.e. the denial of the fgp "for every numeral n, $n + 1 = 1 + n$") cannot be interpreted as an infinite conjunction, and that it is on this account that it is non-finitary in character.

Thus, like his first ground, Smorynski's second ground for introducing the fgp's is mistaken. Hilbert's classification of universal generalizations as meaningful and (unbounded) existential generalizations as meaningless was not due to any belief on his part that the former but not the latter are infinite conjunctions. Hilbert did not take universal generalizations to be infinite conjunctions. Nor did he take interpretability an an infinite conjunction to be a criterion of finitary meaningfulness; indeed, he was of the contrary opinion.

Smorynski's chief mistake is a failure to properly recognize the ways in which Hilbert saw the real propositions as being subdivided. His basic – and most important – distinction was that separating the *problematic* from the *unproblematic* reals. When a finitary proposition or expression is such that one cannot apply the full range of classical logical operations to it without generating a non-finitary proposition, then it was what Hilbert broadly referred to as 'problematic'. But not all problematic finitary propositions are problematic for the same reasons. A bounded existential quantification like "there is a prime greater than **p** but less than **p**! + 1" is a problematic real proposition because it classically implies the unbounded existential quantification "there is a prime greater than **p**", which is not a finitary proposition (cf. Hilbert, 1925, pp. 377–78).

Finitary generalizations, too, are problematic, but for a different reason. When a finitary generalization is negated, one does not get a finitary proposition at all (nor even a finitary proposition-schema). One gets instead an ideal proposition (which is neither a genuine proposition nor a proposition-schema, but rather some sort of formula or computational device). Thus finitary generalizations are problematic because the classically valid law of excluded middle cannot freely be applied to them (cf. Hilbert, 1925, p. 378).[6]

This, in brief, is Hilbert's account of finitary generalizations. It shows what, in Hilbert's view, separates finitary generalizations from other finitary propositions, and why, contrary to what Smorynski suggests, they are to be seen as forming a sub-category of the reals rather than a separate category alongside them. It does not, however, say what is supposed to separate finitary generalizations from ideal generalizations. We must, therefore, briefly address this question.

On Hilbert's view what sets the two apart is the different ways in which they are instantiated. Part of this difference consists in the different character of the instances themselves; or, rather, the differences between the expressions substituted for the variables in producing instantiations of finitary generalizations and ideal generalizations. Hilbert characterized this difference as follows (cf. Hilbert, 1931, p. 194).

Es sei daran errinert, *daß* die Aussage $(x)F(x)$ viel weiter reicht als die Formel $F(z)$, wo x eine beliebig vorgelegte Ziffer ist. Denn im ersteren Falle darf in $F(x)$ für x nicht bloß eine Ziffer, sondern auch ein jeder in unserem Formalismus gebildete Ausdrück vom Zahlcharakter eingesetzt werden . . .[7]

Smorynski presents this as the whole of Hilbert's distinction between finitary and ideal generalizations, saying that the difference between the two is that the former take only numerals (and other finitarily well-defined terms) as substituends of the variables, while the latter also admit "meaningless infinitary constructs" as substituends (cf. Smorynski, 1988, pp. 63–64). But though this is part of the difference between finitary generalizations and ideal propositions, it is not the whole difference, nor, as we shall now attempt to show, even the major difference.

Hilbert repeatedly stressed the fact that there are formulae like '$1 + x = x + 1$' which, though perhaps *interpretable* as finitary generalizations, are nonetheless ideal propositions. One example of this (cf., 1925, pp. 379–80) is the formula '$x + y = y + x$'. This formula, he says, is admissible from the finitist point of view as expressing the generalization that for all (finitary well-defined) numerals m, n, $m + n = n + m$. Yet despite this fact, he says, '$x + y = y + x$' is still an ideal proposition.

His reasons for saying this are important both for understanding the nature of ideal propositions and for seeing what it is that essentially separates them from finitary propositions (or proposition-schemata). Hilbert focused on the question of what *governs the use of* a given formula or proposition; specifically, whether it is the rules of some formal procedure (a 'proof procedure', in his terminology), on the one hand, or consideration of the content of the proposition expressed by the formula in question, on the other. An ideal formula, he says, is not such that its use is guided by appeal to the content of any proposition that the formula in question might (under an assumed semantics) express. Rather it is governed by a given system of rules of 'algebraic' or syntactical manipulation. The use of a finitary formula (or proposition-schema), on the other hand, *is* guided by the considerations of the content (including its evidentness) of the proposition which it (under a given finitary interpretation) expresses.

It follows from this that what determines whether a formula like '$1 + x = x + 1$' is finitary or ideal is not whether, under a specified interpretational scheme, it expresses a finitary proposition or proposition-schema. Nor is it whether the substituends for the variables of the formula are restricted to the class of finitarily well-defined terms. Rather, it is whether the use of the formula is governed by the rules of a formal 'proof procedure', or by considerations of the content of the proposition which it expresses. This seems to be what Hilbert had in mind when he insisted that (cf., 1925, p. 380)

> . . . even when a proposition, so long as it is combined with some indication as to its contentual interpretation, is still admissible from our finitist point of view, as, for example, the proposition that always
>
> $$m + n = n + m$$
>
> where m and n stand for specific numerals, we do not select this form of communication but rather take the formula
>
> $$a + b = b + a.$$
>
> This is no longer an immediate communication of something contentual at all, but a certain formal object, which is related to the original finitary propositions
>
> $$2 + 3 = 3 + 2$$
>
> and
>
> $$5 + 7 = 7 + 5$$
>
> by the fact that, if we substitute numberals 2, 3, 5, and 7, for a and b in that formula (that is, if we employ a proof procedure, albeit a very simple one), we obtain these finitary particular propositions. Thus we arrive at the conception that a, b, $=$, and $+$, as well as the entire formula
>
> $$a + b = b + a$$
>
> do not mean anything in themselves, any more than numerals do. But from that formula we can indeed derive others; to those we ascribe a meaning, by treating them as communications of finitary propositions. If we generalize this conception, mathematics becomes an inventory of formulas – first, formulas to which contentual communications of finitary propositions [hence, in the main, numerical equations and inequalities] correspond and, second, further formulas that mean nothing in themselves and are the *ideal objects of our theory*.[8]

In short, then, a formal proof procedure whose substitutions are restricted to those in which a variable-occurrence is replaced by a numeral or other finitarily well-defined term is, for all that, still a formal proof procedure. And that is why '$1 + x = x + 1$' functions as an ideal formula when one obtains '$1 + 3 = 3 + 1$' from it by appealing to a simple proof procedure which calls for subsitution of the numeral '3' for the variable 'x' (cf. the passage from Hilbert, 1927, quoted in Note 8). If the use of a formula is governed by such a procedure, then it functions (in that context) as an ideal formula, regardless of whether the substituends for its variables are finitarily well-defined. If, on the other hand, a formula like '$1 + x = x + 1$' is used to formalize a piece of reasoning that is based on considerations of the content of the proposition it expresses, then, in Hilbert's view, it functions as a real formula. The critical question is thus whether a formula is used to formalize a proposition which plays a role in a contentual argument, or whether it is used sheerly as an element in a formal 'computational' procedure.

[N.B. Actually, things are more complicated than even the above remarks would indicate. For Hilbert may have assumed that when we employ a formula like '$1 + x = x + 1$' in an ideal way (i.e. as part of a proof procedure), its logic (syntactically speaking) is classical logic. If this is so, then, since classical quantification is quantification *over* objects rather than *objects-as-referred-to-in-some-particular-way*, the substitution class for 'x' would have to be the wider class including terms which are not finitarily well-defined. Perhaps Hilbert had some such assumption in mind; at any rate, it seems to fit with his idea that the point of having ideal methods in mathematics is to preserve the natural and efficient rules of classical logic.]

Hilbert's real/ideal distinction thus consisted of a major division between the real and the ideal propositions, and a minor (sub)division of the real propositions into problematic and unproblematic. In this scheme, the finitary generalizations are merely one particular type of problematic real proposition and not, as in Smorynski's account, a separate class of elements standing alongside the reals and the ideals, nor even a third subcategory of the reals standing alongside the subcategories of problematic and unproblematic. Failing to see this, Smorynski misses the basic point of Hilbert's subdivision of finitary thought; namely, that of separating those finitary elements whose logic abides by the 'natural' principles of classical reasoning (the so-called 'unproblematic' reals) from those (the problematic) whose logic is the more cumbersome and less 'natural' one of finitary reasoning. Missing this basic point, he thus confuses both the general character of Hilbert's conception and its underlying motivation.

III. THE G1-BASED ARGUMENT AGAINST HILBERT'S PROGRAM

With the above discussion of the real/ideal distinction as background, we may now address the main concern of this paper; namely, the G1-based argument against HP. Smorynski (1985, pp. 3–4) presents this argument as follows.

> Hilbert's Programme can be described thus: There are two systems, nowadays called formal theories, S and T of mathematics. S consists of the finite, meaningful statements and methods of proof and T the transfinite, idealized such statements and methods. The goal is to show that, for any meaningful assertion G, if T $\vdash$ G then S $\vdash$ G. Moreover, this is to be shown in the system S.
>
> Gödel destroyed Hilbert's Programme with his First Incompleteness Theorem by which he produced a sentence satisfying a sufficiently narrow criterion of meaningfulness and which, though readily recognized as true – hence a theorem of the transfinite system T, was unprovable in S. In short, he produced a direct counterexample to Hilbert's desired conservation result.

In order to properly evaluate this argument, it will prove useful to give a more explicit version of it. As a step in this direction, we may begin by noting that, in Smorynski's argument, the ideal theory T is treated as representing a *norm* for ideal theorizing; that is, it is taken to be an ideal theory which proves those real sentences that an ideal theory *ought* to prove. These real sentences, on Smorynski's view, are those which are "readily recognized as true". The norm in operation here is thus a type of completeness constraint requiring that an ideal theory be complete w.r.t. those real sentences in its language that are readily recognizable as true.

The assumption that T has this normative status is critically important to the success of Smorynski's argument. This is so because: (i) it is only in its capacity as a prover of the recognizedly true real sentences that T can be said to prove the Gödel sentence G, (ii) it is only in its capacity as a prover of G that T can be shown (by appeal to G1 for S) to prove a real sentence not provable in S,[9] and (iii) it is only because T proves some real sentence not provable in S that it fails to be conservative and hence subject to the sort of defense to which Smorynski takes HP to be committed. The basic thrust of Smorysnki's argument, then, is that an ideal theory is deficient if it fails to prove every recognizable real truth (formulable in its language) as a theorem, and that a deficiency of this sort would be roughly as serious as a violation of conservation. Thus, ultimately, the G1-based argument is intended to present a dilemma to the following effect: the Hilbertian's ideal theories must either fail in their obligation to prove all true real sentences (formulable in their respective languages), or they must fail to be conservative w.r.t. real mathematics. Either way, HP fails.

We find this argument unconvincing for two reasons which we shall now briefly sketch and discuss in greater detail below. The first of these concerns the assumption (hereinafter referred to as *Smorynski-completeness*) that, to be adequate, an ideal theory must prove all recognizedly true real sentences of its language. On an instrumentalist conception of ideal mathematics (which is what we take the conception adopted by the Hilbertian to be), the goal of ideal theorizing is the relatively modest one of proving more efficiently what real theorizing would only prove less efficiently. Specifically, there is no basis in the Hilbertian's instrumentalistic program for requiring that the ideal theory prove *more* real results than the real theory that it is intended to replace. This being so, there is no evident need for the ideal theory T to prove a given sentence unless S also proves it. As shall be argued below, this feature of the Hilbertian's program calls the legitimacy of Smorynski-completeness into doubt.

Our second concern regarding the G1-based argument centers on its demand (henceforth referred to as the *conservation condition*) that ideal mathematics be a conservative extension of real mathematics. The likening of the role of real proof to that of observational verification in the natural sciences suggests that the basic constraint on ideal theorizing is that of a soundness condition requiring that ideal mathematics not prove any real theorem that is refutable by real means. But whereas such a soundness condition naturally gives rise to a corresponding conservation condition in the case of observation statements, it does not do so in the case of real propositions. Hence, the conservation condition is called into doubt.

These then, in outline, are our two main objections to the G1-based argument. We shall develop each in greater detail below. Each, we believe, is sufficient by itself to refute the G1-based argument. However, our concern is not so much to refute that argument as to understand it and its limitations. Since both objections promise to contribute something to this, both would seem to be worthy of the more detailed discussion we now give them.

A. *In order for a theory belonging to ideal mathematics to be adequate, must it be complete w.r.t. the true real sentences formulable in its language?*

In order to answer this question, we must begin by clarifying what is meant by "true real sentence", and our claim is that it is a classical rather than a constructive (specifically, a finitary-constructive) notion of truth that is involved in here. Thus, what Smorynski-completeness demands is that all *classically* true real sentences formulable in L(T) also be provable in T.

That this is the way that Smorynski-completeness should be understood is apparent from the fact that the notion of truth appearing in it must be the same as that according to which G is true, and that notion of truth is the classical one. The intent of Smorynski-completeness is clearly to lay down a requirement on ideal mathematics which T fails to satisfy by failing to prove G. But what makes G true is the truth of its instances, not its provability by finitary means. Hence, it is only classically and not constructively (in particular, not finitarily) true. Consequently, T's 'failure' to prove G can only be counted as a 'failure' to prove all classically true real sentences formulable in its language; which means that Smorynski-completeness must be seen as the requirement that an ideal theory T prove all classically true real sentences formulable in its language.

The significance of this fact for our argument is that it clarifies the possible defenses for Smorynski-completeness as a constraint on ideal theorizing. In particular, it shows us that it cannot merely be seen as an attempt to enforce a simple *strength* requirement on T to the effect that T be powerful enough to codify the whole of finitary reasoning. This is clear from the fact that the theorems of finitary reasoning are not at all the same as the classically true real sentences formulable in L(T), since the latter clearly include sentences that do not belong to the former. What we must now consider is whether there is some other justification for it.

We believe that there is not, and that this is clear from the instrumentalist character of HP. Smorynski-completeness is attractive to a realist, since one of the realist's primary duties in constructing a theory is to bring all truths pertaining to its subject matter under its purview. But the responsibilities of the instrumentalist are different. His task is to replace a less efficient means for identifying a given body of truths with a more efficient one. But that creates no obligation for him to prove anything more than is provable by the more cumbersome methods that he would replace. Once this point is properly appreciated, the inappropriateness of Smorynski-completeness as a criterion of adequacy for ideal theorizing is apparent. Since G is not provable by real means, there is no reason why an ideal theory – whose aim is to improve upon the efficiency of real methods of proof – should be required to prove G either.

B. *Should ideal mathematics be a conservative extension of real mathematics?*

The central idea of the argument of the preceding subsection can readily be extended to provide the starting point for the argument of this one. This extended idea runs as follows: since (i) the Hilbertian's chief obligation is to the efficiency and reliability of his ideal systems, and since (ii) there is no apparent connection between these properties and the ability of an ideal system to decide every real proposition formulable in this language (this being suggested especially by the fact that finitary reasoning does not itself decide every real sentence), it follows that (iii) there is likewise no apparent reason to demand that an ideal theory decide every real proposition formulable in its language. The significance of this conclusion for the present argument is as follows: without a requirement of the finitary decidability of real sentences, the usual conservation condition cannot be derived from the more basic requirement of soundness in the way typical of scientific theories generally.

[N.B. Demanding that an ideal theory T decide the same range of real sentences as is decided by its real counter-part S is not at all the same thing as demanding that T decide all real sentences formulable in L(T). This is so because S itself might not, and indeed typically does not, decide all real sentences formulable in L(T). The Gödel sentence G is a good example. Thus, T may have the scope necessary to serve as a replacement for S (i.e. T may decide every real sentence decidable by S) even if it does not decide every real proposition formulable in its language.]

We begin our argument by noting that the most important feature of S (= real mathematics) is its (presumed) epistemic authority. By this we mean that it is supposed to be the final judge concerning the truth or falsity of real sentences. Its veracity is thus treated as assured, which means that any real proposition it decides must be decided in the same way by any ideal theory which also decides it.

This condition on the real results of ideal reasoning does not, however, imply the conservation condition. It is one thing to say that any real proposition ρ which S decides must be decided in the same way by T, if T decides it, and quite another to say that S must prove every real proposition ρ provable in T. The reason for this divergence is, of course, that S may not decide ρ at all; in which case, respect for the epistemic authority of S provides no reason to demand that ρ be provable in T only if it is also provable in S. To put it another way, when S does not decide ρ, T cannot transgress against the authority of S by deciding ρ. Thus, under such circumstances it is freed from the obligation to adhere to the conservation condition. We conclude, therefore, that the conservation condition is too strong, and should be replaced by the following weaker condition.

(Weak Conservation): For every real sentence r of L(T) such that r is decided by S, if r is provable in T, then r is provable in S.[10]

If one's account of real sentences and finitary reasoning allows that there are real sentences that are not finitarily decideable (i.e. not

S-decidable), then the weaker condition just stated entails no obligation to prove of those real sentences that they are provable in T only if provable in S. Thus, in particular, it does not entail an obligation to prove that G is provable in T only if it is provable in S. Nor, indeed, does it require that this even be true. The unprovability of G in T therefore does not constitute a counter-example to Weak Conservation. Consequently, even if (contrary to the argument of the preceding subsection) T were required to prove G in order to be an adequate ideal theory, G's unprovability in S would not constitute a violation of the more appropriate condition of Weak Conservation.

To allay the suspicions of those who are drawn to the usual conservation condition, it is necessary to offer some explanation of why, despite its ultimate indefensibility, the ordinary conservation condition should nonetheless prove so alluring. Our attempt at doing so begins with a very basic tenet of Hilbert's Program; namely, the alleged analogy, mentioned earlier, between the epistemic roles played by observation in physics and real proof in mathematics. This analogy suggests, albeit falsely, that if conservation w.r.t. observational consequences is a reasonable condition to place on a physical theory,[11] then conservation w.r.t. real consequences ought to be a reasonable condition to place on an ideal mathematical theory. Thus, the reasoning continues, if it can be shown that conservation w.r.t. observational consequences is a reasonable condition to place on physical theorizing, it should follow that conservation w.r.t. real consequences is a reasonable condition to place on the Hilbertian's ideal theorizing.

One can, of course, argue quite convincingly that conservation w.r.t. observational consequences is a reasonable condition to place on physical theorizing. The argument begins by noting that in physics (as in natural science generally) observation is granted a place of special epistemic privelege; by which it is meant that the theoretical elements of physics are not at liberty to oppose that which is established by observational means. This epistemic subordination of theoretical proof to observational verification is expressed as a principle of *observational soundness*, which is then taken to constitute a condition of adequacy on any body T of theoretical reasoning in physics:

(OS) For any observation sentence O, if T proves O, then O cannot be observationally falsified.

From (OS) as a starting point, it is possible to argue for the following principle of observational conservation

(OC) For any observation sentence O, if T proves O, then O is verifiable by observational means,

by adding to (OS) the further premise that observation sentences are (by their very nature?) supposed to be decidable by observational evidence. If, as (OS) requires, no observational consequence O of T can be observationally refuted, and every observation sentence is observationally decidable, then every observational consequence of T must be observationally provable. Hence, (OC) follows from (OS) and inherits its plausibility.

Can a similar defense for a condition of real-conservation on ideal theorizing be given? A condition of real soundness would appear to be no less defensible a constraint on ideal theorizing than (OS) is on physical theorizing:

(RS) For any real sentence r, if T proves r, then r cannot be refuted by real means (i.e. $\neg r$ is not probable in S).

But in order to get a condition of *real conservation* from (RS), one requires an additional premise stating that every real sentence is decidable in S; and such a premise is false – at least if one counts such sentences as G as real.[12] It is therefore impossible to base a case for real conservation on the presumed need for real soundness in the same way that a case for observational completeness can be based on a presumption of need for observational soundness.

We conclude our discussion of the conservation condition by briefly considering a possible objection to the argument just given. This objection focuses on that feature of real sentences which is responsible for the above-noted separation of real-conservation from real-soundness; namely, their undecidability by real means. It has, moreover, both a conceptual and an historical side to it. On the conceptual side, it holds to a constructivistic understanding of real propositions (including those generalizations which Hilbert referred to as 'hypothetical judgements'), and maintains that such a view demands that genuinely real propositions be decidable by finitary means. On the historical side, it would maintain that Hilbert was originally committed to the finitary decidability of real propositions, and present G1 as having refuted that belief. Hence, it views Hilbert's original program as having based a commitment to real conservation on a deeper commitment to real soundness, and therefore sees G1 as destroying that hope.

On the textual-historical side, there is the well-known remark in the 'Mathematical Problems' address claiming that ". . . every definite mathematical problem must necessarily be susceptible of an exact settlement, either in the form of an actual answer to the question asked, or by a proof of the impossibility of its solution and therewith the necessary failure of all attempts [to solve it]" (1901, p. 444, brackets mine). This same theme was sounded in Hilbert (1925), in the famous claim stating that in mathematics there is no *ignoramibus*. If 'definite problems' are just real propositions in interrogative form, it is hard to see how Hilbert could have counted undecidability proofs as settling genuine problems while also holding the position that every real proposition must be finitarily decidable. Since propositions stating the undecidability of a given proposition by various means were held by Hilbert to be genuine propositions, this suggests that he would not have accepted the view that all genuine (= real) propositions must be finitarily decidable.

These textual points aside, however, it is doubtful that a constructivist semantics for real propositions *should* equate meaningfulness with finitary decidability. On a constructivist account, a user of the language is said to know the meaning of a real sentence r when for every proof (resp. refutation) π of r, she *would* recognize π as such *were* she to be presented with it. This does not imply, however, that real (i.e. meaningful) sentences are finitarily decidable – not even if it is assumed that the only kinds of proofs of refutations of real sentences there are, are finitary. That this is so follows from the fact that being in a position to recognize a proof or refutation of r were one to be presented with it does not require knowing how to prove or refute it. Ability to tell of a given proof whether or not it is a proof or refutation of r does not imply ability to actually generate such a proof or refutation. There is thus no reason to maintain that, on a constructivistic account of their meaning, real sentences must be finitarily decidable.

Since, then, there are textual reasons for denying that Hilbert held the view that real propositions are finitarily decidable, and conceptual reasons against such a view regardless of its textual pedigree, it seems only charitable to withhold attributing such a view to Hilbert. We therefore reject the suggestion that he was committed to the finitary decidability of real propositions and that his commitment to real-soundness should therefore be seen as engendering a commitment to real-conservation.[13]

IV. CONCLUSIONS

The argument of the preceding section defeats the G1-based argument against HP. In so doing, it re-focuses attention on the G2-based argument. This, we believe, is all to the good, since it calls attention to those problems which represent the deepest philosophical issues concerning Gödel's Theorems and their implications for Hilbert's Program; issues on which the literature – including the philosophical literature – has been strangely silent. At bottom, the central question is "What is an instrumentalist theory?" We shall argue that the proof of G2 (in particular, the familiar Derivability Conditions) places constraints on what is to count as an ideal theory that are unwarrantedly strict. Moreover, we shall argue that the G2-based argument against HP fails to properly distinguish two very different concerns regarding instrumentalist theories; namely, (i) whether they prove all they need to prove in order to be adequate replacements for the real theories that they are supposed to replace, and (ii) whether they prove only such real theorems as are not refutable by real means. Specifically, it fails to take proper account of the fact that the Hilbertian instrumentalist requires a finitary proof only of (ii) and not of (i), and that the two should therefore not be merged into a single condition to be finitarily proven.

Once a suitably liberal standard of what is to count as an instrumentalist theory is adopted, and once it is realized that though the

Hilbertian may have need of a finitary proof of the real-soundness of his ideal theories, he has no similar need for finitary assurance that they prove all of what he wants them to prove, the force of the G2-based argument is dissipated. Such, at any rate, is our thesis, though our basic aim is not so much to establish it as to call attention to those deeper philosophical issues from which it arises.[14]

These issues can perhaps best be got at by considering the differences between the conditions which the proofs of G1 and G2 place on their underlying notions of theory. The proof of G1 makes only minimal demands. It is concerned only with the content of a theory; i.e. with *what* it proves. Specifically, it demands only that the theory in question contain enough recursive number theory to weakly represent the set of (Gödel numbers of) its theorems.[15] One might therefore say that the notion of theory presupposed by the proof of G1 is one which identifies a theory with the set of theorems that it proves, and whose only additional constraint is that it contain enough recursive number theory to represent that set.

G2, on the other hand, makes more extensive demands on the notion of theory. Its proof depends not only on a theory's having the right content, but also on that content's having been admitted as such by a certain type of procedure. Two theories T_1 and T_2 may contain exactly the same entities in a given metamathematical category (e.g. the category of formulae, axioms, proofs, or theorems) and yet differ significantly with respect to the conditions used to qualify items for membership in it.

To illustrate what we are talking about, let us recall the examples of consistency-minded theories mentioned in the introduction and how they contrast with the standard conception of a formal axiomatic theory. On the standard conception, the basic metamathematical categories pertaining to a theory (e.g. those of formulae and axioms) are defined inductively, and the remaining categories (e.g. that of the theorems) are defined in terms of these. One begins the definition of a basic category by exhibiting a finite set of basic members, and then identifying the remaining members with those items that can be generated from the basic members (perhaps taken in combination) by applying (in some instances iteratedly) certain specified syntactical operations to them. The application of these operations is, moreover, supposed to be self-contained in a certain sense. Specifically, it is supposed that in order to determine of a given item whether it belongs to the category in question, one need only attend to the syntactical traits of *its* constituents and their arrangement and not (even potentially) to the syntactical characteristics of other items.

Consistency-minded theories are different. They begin with metamathematical definitions of the standard sort, but then add various conditions which, broadly speaking, are consistency conditions of one or another sort. Thus, to take the Rosserian variant of consistency-minded theories as an example, one begins with a theory T, defined in the standard way, and forms the Rosser variant T_R of it by adding to the standard definition of proof a constraint to the effect that the last line of a proof-in-T_R not be the contradiction of the last line of any proof-in-T which precedes it in a given omega ordering of the proofs-in-T.[16] Thus, the category of proofs-in-T_R is generated by a different testing or qualifying procedure than the category of proofs-in-T, and that is what we mean by saying that the category of proofs-in-T is generated or governed by a different admission procedure than the (possibly coextensive) category of proofs-in-T_R.

Such differences separating the admission procedures for the various metamathematical categories of T from those of T_R will not have any effect on whether G1 holds for them. It will either hold of both or of neither, depending on whether T proves enough number theory to weakly represent the class of (Gödel numbers of) its theorems. Such differences can, however, have a decisive effect on whether G2 applies equally to both T and T_R, since the proof for G2 is sensitive to the character of the admission procedures for the various metamathematical categories pertaining to a given theory.

What this means is that G2 applies not to sets of theorems (or even sets of proofs!) but rather to sets of theorems-as-admitted-by-a-particular-type-of-procedure. This raises the possibility – of crucial importance to the present discussion – that there might be non-standard ways of generating sets of theorems and proofs that are not covered by the proof of G2, but which nonetheless constitute perfectly sensible conceptions of theory for the Hilbertian instrumentalist. Were this to be so, the Hilbertian would be free to construct ideal theories not ruled out by G2, and having the same content as those that are. At any rate, this would seem to be a possibility worth looking into.

These last remarks point out why it so important to get clear about the success of the G1-based argument against HP. For if the G1-based argument were correct, then the prohibition against the Hilbertian would be one bounding the strength or content of his ideal theories (i.e. one fixing limits on what his ideal theories could be capable of proving), and not just one restricting the mode according to which that content is to be generated. If, on the other hand, the G1-based argument is not successful, and the basis of evaluation for HP must accordingly be shifted to G2, then the focus of concern is not one concerning strength or content, but rather one concerning mode of generation (with respect to which the Hilbertian may well have flexibilities that he does not have with respect to the strength of his ideal theories). We regard this as a difference of potentially great significant, and one which makes the investigation of consistency-minded modes of theorizing imperative.[17]

Perhaps the most basic issue raised by the possibility of consistency-minded theorizing is this: Are theories to be viewed *extensionally* (i.e. as sets of beliefs) or *intensionally* (i.e. as methods or procedures for selecting beliefs)? The good theory, of course, both targets particular propositions for belief, and does so by bringing them under a method which testifies to their belief-worthiness. Let us refer to the first of these two components of good theorizing as the *locative component*, and the second as the *methodological component*.

Neither component is, by itself, sufficient for good theorizing. A good theory is not a mere 'list' of propositions to believe, with no credential provided to attest to their belief-worthiness. Likewise, a good theory is more than sheer method, with no means given for finding a set of particular propositions that are to be believed[18]

All this may seem so elementary as to scarcely be worth mentioning. Why then do we emphasize it so? The answer is that it points to an important yet easily overlooked truth concerning the Hilbertian's project; namely, that the locative and methodological features of theorizing are *separable*. In order to defend a given body of ideal theorizing, the Hilbertian must know something about both its locative and its methodological features. But what he must know about them – and this is the important point – is different. Regarding the methodological element, what he must know is that it is real-sound (i.e. that all of the real sentences it decides are decided in the same way by finitary means). Furthermore, this must be made apparent by finitary means if his use of ideal method is to avoid the threat of 'diluting' the knowledge that he might otherwise obtain by foreswearing the use of ideal methods and sticking to (the presumably less efficient) real methods. A gain in efficiency is not so attractive if it brings with it a corresponding drop in the epistemic quality of the more efficiently attained 'knowledge'. The Hilbertian thus proposes to replace the object-level real proofs of a given real sentence ρ with a meta-level real proof of ρ that is comprised of two elements: (i) a real metamathematical proof that ρ is provable in the ideal system T (which would consist in producing a proof of ρ in T), and (ii) a real metamathematical proof showing of ρ that if it is provable in T, then it is also provable by real means at the object-level.[19]

Having this sort of control over the quality of his ideal methods assures the Hilbertian instrumentalist that their use will not engender 'dilution' (i.e. will not result in an epistemic product of a quality lower than what could have been attained by sticking to the real methods that the ideal methods in question are intended to replace), and this is what he needs to know.[20] What the Hilbertian demands of the locative component of a given ideal theory T is that it replicate the locative capacity of that body R of less efficient real reasoning which it is intended to replace. In other words, the Hilbertian must be able to show that every sentence provable in R is also provable in T.[21] However, the Hilbertian has no need for *finitary assurance* that a proposed ideal replacement extensionally simulates the real reasoning it is to replace. Hence, he is under no obligation to prove finitarily that each theorem of R is a theorem of T.

This is significant because though the move from a standard ideal theory T to one of its consistency-minded variants T_Δ generally affords a greater measure of control over the quality of its real theorems, this increase in methodological control brings with it a consequent loss in ability to prove (by means codifiable in T, and hence, typically, by finitary means) the extensional equivalence of T and T_Δ.[22] But, as the reasoning of the preceding paragraph suggests, this 'loss' is not disabling for the Hilbertian. His primary obligation with regard to the content or locative component of T_Δ is to show that it replicates R, not T. And even if he were only able to do this by first proving the extensional equivalence of T_Δ and T, all that would follow from his inability to prove this latter fact finitarily is his subsequent inability to *finitarily* show that T_Δ replicates R. But this is no impediment to his program since what he needs a *finitary* proof of is not the replication of R by T_Δ, but rather the Weak Conservation w.r.t. R of T_Δ. With respect to the replication of R by T_Δ, all he needs is convincing evidence, not finitary proof.[23]

Thus, the general fact that one cannot finitarily prove the theorem-wise equivalence of consistency-minded theories and their standard counterparts gives no reason why the Hilbertian should not use consistency-minded construction techniques for his ideal theories. It poses no obstacle to his getting the kind of control over the locative element of theorizing that he needs, and, since the usual techniques are susceptible to G2 in a way that the consistency-minded techniques are not, it also affords him an advantage over the standard techniques when it comes to managing the methodological factor. This does not, however, show – what is a very difficult question, and one which we are currently not in a position to resolve[24] – that moving to consistency-minded techniques of theory-construction will actually allow one to carry out HP. Nonetheless, it does suggest that part of the traditionally pessimistic view of the Hilbertian's prospects is due to an unwarrantedly narrow conception of how ideal theories should be constructed; a conception which is built into the very fabric of the Derivability Conditions governing G2, and which systematically ignores the possible benefits to the Hilbertian of adopting consistency-minded modes of ideal theorizing.

ACKNOWLEDGEMENT

The author wishes to thank the Alexander Von Humboldt Foundation for its generous financial support.

NOTES

[1] Cf. Hilbert (1927), p. 475 for an explicit statement of the analogy between phyiscal theory and ideal mathematics. Smorynski too (cf., 1988, p. 59) adopts this view. Thus he quotes the remark from Hilbert (1925) that the science of mathematics is not reducible to its real elements, but that it must always yield correct real results, and then goes on to say that one can see in this remark the view that "Mathematics is an abstract theoretical science subject to numerical control just as physics is an abstract theoretical science subject to experimental control."

[2] Throughout this paper, when we speak of an ideal proof we shall generally mean an ideal proof of a real sentence.

[3] In noting that, on the instrumentalist conception, ideal proofs 'yield' (via evaluation in a real metamathematical scheme), rather than constitute, justifications, the question is naturally raised as to why they should be of any epistemic interest. Why, that is, should one be interested in producing justification through the indirect means of ideal proof rather than through the direct means of real proof? Such questions, however, apply as well to the realist conception; for one can as well ask "Why should we be interested in pursuing justification for observation statements through the indirect means of theoretical arguments rather than through the more direct means of observation?".

In the empirical sciences, the question is perhaps a little easier to answer than in mathematics. For there, at least sometimes, the whole idea is to not be in a position to observationally settle a question (e.g. when it is the empirical effects at ground-zero of a nuclear explosion that are in question). Other times, it might not be disutility, but rather possibility, feasibility, cost, and/or inconvenience that would make an alternative to observational justification attractive. In any event, there must be something about justification via use of ideal methods (be they instrumentalistically or realistically conceived) that makes it attractive as an alternative to real justification. In Hilbert's case, this has to do with what might broadly be termed 'efficiency'. On his view, there is a difference between the laws according to which our minds operate most efficiently (viz. the laws of classical logic), and the laws according to which finitary (i.e. real) truth works (i.e. finitistic logic). This is the epistemological predicament of the human who seeks mathematical knowledge. Hilbert's project was to provide a way out of the predicament by showing that we can enjoy the benefits of efficiency of (ideal) classical reasoning without sacrificing the accuracy of (real) finitary reasoning.

[4] There are those (e.g. Gentzen, 1936, 1938; and Prawitz, 1981) who have opted for a more realistic interpretation.

[5] On p. 64, Smorynski says that, ignoring subtleties, ". . . we can say that Hilbert said the following: Let S be a formal system of finitary arithmetic and let T be some system of transfinite mathematics. Suppose S proves the consistency of T . . . Then: For any universal assertion G, if T $\vdash$ G then S $\vdash$ G." This latter is the conservation condition referred to in the text. It makes finitary general propositions function like data or controls on ideal (i.e. transfinite) mathematics, because it forces every finitary general proposition proven in the ideal theory to be corroborated by finitary means.

[6] The interested reader may consult Detlefsen [1986], chs. I and II for further discussion of Hilbert's views of problematic reals (and some of the difficulties associated with them).

[7] My translation: "It should be remembered here that the proposition $(x)F(x)$ extends much farther than the formula $F(z)$, where z may be any one of the specified

numerals. For in the first place, one is permitted to substitute for x in $F(x)$ not just a numeral, but any one of the expressions in our formalism that is constructed from numerical terms . . .".

[N.B. Here '$(x)F(x)$' plays the part of an ideal generalization, and 'Fz' the part of a finitary generalization.]

[8] In this quotation, I have used the Latin characters 'm', 'n', where Hilbert used Gothic characters. The same general view is presented in Hilbert (1927), pp. 469–70, and there his example is exactly the same formula that Smorynski uses (though he writes it as '$1 + a = a + 1$' instead of '$1 + x = x + 1$').

". . . algebra already goes considerably beyond contentual number theory. Even the formula

$$1 + a = a + 1,$$

for example, in which a is a genuine number-theoretic variable, in algebra no longer merely imparts information about something contentual but is a certain formal object, a provable formula, which in itself means nothing and whose proof cannot be based on content but requires appeal to the induction axiom.

The formulas

$$1 + 3 = 3 + 1 \text{ and } 1 + 7 = 7 + 1,$$

which can be verified by contentual considerations, can be obtained from the algebraic formula above only by a proof procedure, such as formal substitution of the numerals 3 and 7 for a, that is, by the use of a rule of substitution."

[9] Since Smorynski's argument depends upon T's proving G, it is clear that G cannot be the Gödel sentence for T. What may be less clear is what theory (or theories) it is of which G is supposed to be the Gödel sentence. As we shall see in a moment, the key constraints on G are that it be (classically) true and not provable in S. Thus, if we take S to stand for a formalization of finitary number theory, G might be taken to be the Gödel sentence of S. For taken in that way, S is true and hence consistent. Hence, its Gödel sentence is not provable in it. Hence, its Gödel sentence is (classically) true. This makes it clear, however, that G might just as well be taken to be the Gödel sentence of any extension of finitary number theory that is clearly consistent.

[10] Stated a little more formally, this conservation condition reads as follows (where T is the ideal system whose conservation is to be proven, S the formalism representing finitary reasoning, and '$\vdash_T$' and '$\vdash_s$' stand for provability in T and S respectively):

For any real sentence ρ of L(T) such that ρ is decidable in S,
$\vdash_S [\vdash_T \rho \rightarrow \vdash_S \rho]$.

[11] Here, in speaking of a physical theory that is conservative w.r.t. its observational consequences, we mean a physical theory which is such that every observation statement it implies is also verifiable by observational means.

[12] Nor is a gambit like Smorynski's distinction between real and finitary general propositions of any use here. What Smorynksi (1988) calls 'real sentences' are all decidable by finitary means. What he classifies as 'finitary general propositions', however, are not, since G is treated as a finitary general proposition. However, the conservation condition is supposed to apply to finitary general propositions (cf. the formulation of

conservation on p. 64 of Smorynski, 1988) as well as to real propositions. Such a conservation condition is no more derivable from a corresponding soundness principle (i.e. a principle of soundness extending to Smorynski's finitary general propositions as well as to what he calls the real propositions) than (RC) is derivable from (RS). Hence, our argument would not be undone by employing Smorynksi's scheme of distinctions.

[13] This, of course, implies neither that Hilbert did not hold that the real sentences were, as a matter of contigent fact, finitarily decidable (though, as indicated by the earlier textual remarks, it is questionable that he did), nor that he was not surprised by Gödel's proof. It is only to say that such a belief, if held at all by Hilbert, was nonetheless extraneous to his program and not one of its essential tenets.

[14] Detlefsen (1986, pp. 120–24) contains the only sustained argument for the adequacy of consistency-minded theories (in particular, Rosser variants) as models of instrumentalist theorizing. Aside from that, the only indications of awareness of the possible inappropriateness of the view of theory that is presupposed by the Derivability Conditions are in Kreisel and Takeuti (1974), Jeroslow (1975), and Kreisel (1980). The former (pp. 47–8) write that the consistency-minded variant of Rosser provides "a neat model for Wittgenstein's speculations" on the nature of rule-following. Jeroslow treats the consistency-minded theories of Feferman (1960) as based on a trial-and-error conception of theory-construction. Finally, Kreisel (1980, p. 173) contains the statement that Rosser variants "mirror quite well, albeit crudely, an essential method used in practice for checking proofs: comparison with background knowledge . . .".

[15] A set of numbers Σ is said to be weakly representable in T if there is a formula σ of L(T) of one free variable such that for every natural number n, $n \in \Sigma$ iff $\vdash_T \sigma(\mathbf{n})$ (where '$\mathbf{n}$' is the standard term in L(T) for n).

[16] When we speak of the Rosser system T_R 'starting with' a standardly defined system T, we are supposing that the standardly defined system T serves as a specification of a body of standard ideal derivations whose advantages we should somehow like to incorporate into our consistency-minded replacement T_R of R. It should thus be clear that there is nothing sacrosanct about building consistency-minded theories from standard ones, and that we would only do so in those cases where we were convinced of their utility as ideal instruments.

[17] How serious are the bounds on the strength of the Hilbertian's ideal theories that are induced by Gödel's work? The so-called 'reverse mathematics' of Friedman and Simpson attempts to show that it is not as bad as might be thought. They have proven that there are systems embodying a substantial portion of classical mathematics all of whose Π_2 theorems are provable in PRA; and Sieg (1985) has subsequently shown how to prove this by finitary means. The idea behind reverse mathematics (and, presumably, the inspiration for its name) is to begin with a particular body of results believed to form the core of classical mathematics (and hence to be an indispensable part of the ideal systems of the Hilbertian), and thence to find the weakest possible set of axioms for proving these. In that way, one ends up with a set of axioms that is (at least conceptually speaking) equivalent to the theorems to be proven, rather than simply strong enough to entail them (and perhaps much stronger than what is really needed). One thus stands to eliminate the unnecessary strength present in the usual axiomatizations, and this may in turn enable him to more nearly approximate Hilbert's goal of a finitary soundness proof for ideal mathematics.

[N.B. I am not so sure, however, that the notion of 'necessary strength' tacitly employed by the advocates of reverse mathematics is the correct one. Simpson, at least, thinks that Hilbert was simply out to save classical mathematics. Hence, he takes the starting point of Hilbert's Program (and thus of reverse mathematics) to be some body of classical results, which must be preserved. I, however, regard this as incorrect. Hilbert did want to preserve classical mathematics, but this was not for him an end in itself. What he valued in classical mathematics was its efficiency (including its psychological naturalness) as a means of locating the truths of real or finitary mathematics. Hence, any alternative to classical mathematics having the same benefits of efficiency would presumably have been equally welcome to Hilbert. There may, of course, be no such alternatives, and it may even be that Hilbert believed this. But regardless of whether this is true (and who knows whether it is), it would still be misleading to describe Hilbert's goal as that of preserving classical mathematics.]

Another strategy for getting a more accurate (and conservative) estimate of the Hilbertian's ideal commitments may be found in Detlefsen (1986), chs. III and V. The idea there is also a sort of 'reversing' strategy, however one which starts not from what is taken to be essential to classical mathematics (qua mathematics) but rather from what is judged to be the instrumentally useable portion of ideal mathematics. It begins with the belief that only some of the proofs in the usual ideal systems will be efficient enough to prove any advantage in efficiency to the Hilbertian instrumentalist; others will either be too long to be useable as an instrument, or they will not afford any advantage in efficiency over their real counterparts. In principle, such proofs could be eliminated from the ideal systems without any cost to the instrumentalist (i.e. without any loss in the efficiency of the ideal system). The fragment of the system remaining after such eliminations is what the instrumentalist is really responsible for, and his soundness proofs ought therefore ideally be aimed at them rather than the usual systems. In essence, then, the idea is to start with the instrumentally gainful ideal proofs and work backward to a minimal system capturing them.

Like the 'reverse mathematics' of Friedman and Simpson, this approach to Hilbert's Program attempts to revive it by being more exact (and conservative) about the strength required for ideal mathematics. This connection on the strength of the theories needed by the Hilbertian should be contrasted with our present emphasis, which is to raise the question of what happens to Hilbert's Program when one leaves the strength of the usual ideal theories intact, but alters the mode according to which they are generated. It may be, however, that the best way to develop Hilbert's Program would be to combine elements of both sorts of approaches; that is, to modify both the strength of the Hilbertian's ideal commitments and their mode of generation.

[18] The locative and methodological elements may not, of course, be so cleanly defined. In particular, the locative element of a theory may also play a role in its methodological control over beliefs. Generally speaking, this will happen when there is a means of judging the quality of particular outputs of the locative element that is independent of that sponsored by the general description we have of the quality of its outputs. To take an example, we might believe of a set of well-confirmed empirical generalizations that they are true and that the logic we use to derive observational consequences from them is truth-preserving. Still, we have an independent means (viz. observation) of evaluating those consequences. And if that independent means sponsors a contrary judgement, it might even cause us to reassess the judgement of quality based on the general qualitative description we have of the locative element (viz. that the generalizations concerned are true, and/or the logic for manipulating them truth-preserving).

When part of our access to the epistemic quality of the locative element is of this "consequentialist" type, the general description we have of its epistemic quality will not form the whole of its methodoligcal element.

But though this is true and may even be typical of theories, it is important not to lose sight of the need for a description of the locative element (i.e. a methodology) that gives a general assessment of its output. For the whole idea behind a theory typically is that we either cannot or do not want to try to gain independent access (i.e. access not provided by following the prescribed method of belief-selection) to the outputs of the locative element. Thus, for example, an empirical theory is desirable just because we do not have timely and/or sufficiently safe and economical observational access to its observational consequences. Likewise, for the Hilbertian, ideal mathematics is desirable precisely because we do not have suitably efficient access to the truths of mathematics via real proof. Thus, having independent access to the epistemic quality of the outputs of the locative element of a theory does not eliminate the need for a *general* assessment of its outputs.

[19] Where does the proof referred to in (ii) come from? In particular, what sorts of soundness principle does it come from? In the last section, we argued that the Hilbertian is not committed to the usual conservation condition since he is not responsible for proving of real sentences that are finitarily undecidable that they provable by ideal means only if they are also provable by real (i.e. finitary) means. Thus, he need not show of each real sentence formulable in the language of a given ideal theory T that it is provable in T only if it also provable by real means. This restricted soundness principle can, however, only give rise to a proof of the sort referred to in (ii) when it can also be shown that the real sentence for which one has an ideal proof (by (i)) is decidable by real means, and this is something the ideal theorizer would like to avoid having to show (since he has no general finitary procedure for doing so).

There is, however, a more congenial alternative modification of the usual conservation condition which would seem to be available. The idea behind this alternative is that the Hilbertian has no need to show of real sentences that are *not* decidable by the *ideal* theory he is defending that they are provable in it only if they are also provable by real means. This is so because he will most assuredly not make use of any ideal proofs of such sentences; and if he cannot make use of them, there is no reason why they should be included within the scope of his soundness condition. Modified accordingly, the Hilbertian's obligation reduces to that of showing (finitarily), of each real sentence decidable in the ideal theory he is defending, that it is provable in that theory only if it is provable by real means. Such a restricted soundness condition, in concert with proof referred to in clause (i), affords the Hilbertian a proof of the sort referred to in (ii).

[20] For a refinement and extension of these brief remarks concerning the threat of epistemic 'dilution' and its place in the Hilbertian's thinking, see Detlefsen (1986), chs. I and II.

[21] Note that here R need not be identified with finitary reasoning *per se*. It is only that body of finitary reasoning that is to be replaced by the ideal system T, and this might not constitute the whole of finitary reasoning. Indeed, the Hilbertian is not committed to holding that the whole of finitary reasoning ought to be replaced; there might, after all, be cases of finitary reasoning whose efficiency cannot be improved upon by any known ideal means. These possible differences between R and the whole body of finitary reasoning also indicate that it would be a mistake to demand (by

way of a soundness principle) of the real sentences provable in T that they should be provable in R.

[22] To show this, one appeals to (a) G2 for T, (b) a premise to the effect that if the theorem-wise coextensivity of T and T_Δ were provable in T, then 'Con (T_Δ) $\rightarrow$ Con (T)' would also be provable in T, and (c) the fact that Con (T_Δ) is provable in T.

[23] Note that even with standard (as opposed to consistency-minded) theories, one does not have finitary control over the locative element, since once cannot show finitarily that they do not prove some absurdity.

[24] Hilbert's Program, as I understand it, only requires a proof of the real-soundness of the *instrumentally gainful* ideal methods (i.e. those ideal derivations which are (i) short and simple enough to be of some use by agents with our cognitive limitations (possibly with the assistance of realizable computing machines), and (ii) more efficient than any available real proof of the same result). This, however, requires that we have some sort of 'complexity metric' for rating (and comparing) the complexity of real and ideal proofs. There are, of course, various complexity metrics that have found their way into the proof-theoretic literature, and the recent literature in theoretical computer science has produced even more. Yet all of these complexity metrics seem to be designed to measure a general type of complexity that might be called 'verificational complexity'; that is, the type of complexity that is encountered in determining of a given syntactical entity whether or not it is a proof in a given system of proofs. It seems, however, that what the Hilbertian is chiefly concerned with is not verificational, but rather 'inventional' complexity; that is, the type of complexity that is encountered in coming up with a proof in the first place (as opposed to verifying of a *given* item that it is a proof). This is strongly suggested by Hilbert's statement (1927, p. 475) that his ". . . formula game is carried out according to definite rules, in which the *technique of our thinking* is expressed. These rules form a closed system that can be discovered and definitively stated. The fundamental idea of my proof theory is none other than to describe the activity of understanding, to make a protocol of the rules according to which our thinking actually proceeds. Thinking, it so happens, parallels speaking and writing: we form statements and place them one behind another. If any totality of observations deserves to be made the object of a serious and thorough investigation, it is this one."

If this way of looking at Hilbert's Program is right, then, in order to properly evaluate it, a metric for *inventional* complexity would have to be developed. Since, however, we are far from being able to do this, we are equally far from being able to give a definitive evaluation of Hilbert's Program . . . the traditional negative arguments and more recent positive proposals (e.g. reverse mathematics) notwithstanding. Much basic philosophical work remains to be done.

REFERENCES

Arai, T.: 1990, 'Derivability Conditions on Rosser's Provability Predicates', to appear in the *Notre Dame Journal of Formal Logic*.

Detlefsen, M.: 1986, *Hilbert's Program*, D. Reidel Publishing Co., Dordrecht.

Feferman, S.: 1960, 'The Arithmetization of Metamathematics in a General Setting', *Fundamenta Mathematicae* **49**, 35–92.

Gentzen, G.: 'The Consistency of Elementary Number Theory', in M. E. Szabo (trans. and ed.), *The Collected Works of Gerhard Gentzen*, North-Holland Publishing Co., Amsterdam, 1969.

Gentzen, G.: 1938, 'The Present State of Research into the Foundations of Mathematics', in M. E. Szabo (trans. and ed.), *The Collected Works of Gerhard Gentzen*, North-Holland Publishing Co., Amsterdam, 1969.
Guaspari, D. and R. Solovay: 1979, 'Rosser Sentences', *Annals of Mathematical Logic* **16**, 81–99.
Hilbert, D.: 1901, 'Mathematical Problems', *Bulletin of the American Mathematical Society* **8**, 437–479.
Hilbert, D.: 1925, 'On the Infinite', in J. van Heijenoort (ed.), *From Frege to Gödel*, Harvard Univ. Press, Cambridge, 1967.
Hilbert, D.: 1927, 'The Foundations of Mathematics', in J. van Heijenoort (ed.), *From Frege to Gödel*, Harvard Univ. Press, Cambridge, 1967.
Hilbert, D.: 1930, *Grundlagen der Geometrie*, Teubner, Leipzig and Berlin, 7th. ed.
Hilbert, D.: 1931, 'Die Grundlegung der elementaren Zahlenlehre', in *Gesammelte Abhandlungen*, vol. 3, Julius Springer-Verlag (1935), Berlin.
Jeroslow, R.: 1975, 'Experimental Logics and Δ_2^0-Theories', *Journal of Philosophical Logic* **4**, 253–267.
Kreisel, G.: 1971, 'A Survey of Proof Theory II', in J. E. Fenstad (ed.), *Proceedings of the Second Scandinavian Logic Symposium*, North-Holland Publishing Co., Amsterdam.
Kreisel, G. and G. Takeuti: 1974, 'Formally Self-Referential Propositions for Cut-Free Analysis and Related Systems', *Dissertationes Mathematicae* **118**, 4–50.
Kreisel, G.: 1976, 'What Have We Learnt from Hilbert's Second Problem?', *AMS Proceedings of Symposia in Pure Mathematics* **28**, American Mathematical Society, Providence.
Kreisel, G.: 1980, 'Kurt Gödel', *Biographical Memoirs of the Fellows of the Royal Society* **26**, 149–224.
Prawitz, D.: 1981, 'Philosophical Aspects of Proof Theory', in *Contemporary Philosophy: A New Survey*, vol. 1, Martinus Nijhof Publishers, The Hague.
Rosser, J. B.: 1936, 'Extensions of Some Theorems of Gödel and Church', *Journal of Symbolic Logic* **1**, 87–91.
Sieg, W.: 1985, 'Fragments of Arithmetic', *Annals of Pure and Applied Logic* **28**, 33–71.
Simpson, S.: 1988, 'Partial Realizations of Hilbert's Program', *Journal of Symbolic Logic* **53**, 349–363.
Smorynski, C.: 1977, 'The Incompleteness Theorems', in J. Barwise (ed.), *Handbook of Mathematical Logic*, North-Holland Publishing Co., Amsterdam.
Smorynski, C.: 1985, *Self-Reference and Modal Logic*, Springer-Verlag, New York.
Smorynski, C.: 1988, 'Hilbert's Programme', *CWI Quarterly* **1**, 3–59.
Visser, A.: 1989, 'Peano's Smart Children: A Provability Logical Study of Systems with Built-in Consistency', *Notre Dame Journal of Formal Logic* **30**, 161–196.

Department of Philosophy,
University of Notre Dame,
Notre Dame, IN 46556,
U.S.A.