Finitism:
An Essay on Hilbert's Programme

by

David Watson Galloway

BMus, University of Wales (1972)
MPhil, University of London (1989)

Submitted to the Department of
Linguistics and Philosophy
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

at the Massachusetts Institute of Technology

May 1991

Signature of the Author .. ..
David Galloway
Department of Linguistics and Philosophy
May 1st 1991

Certified by..... .....
George S Boolos
Professor, Department of Linguistics and Philosophy
Thesis Supervisor

Accepted by..... .....
George S Boolos, Chairman
Departmental Graduate Committee
Department of Linguistics and Philosophy

Finitism:
An Essay on Hilbert's Programme

by

David Watson Galloway

Submitted to the Department of Linguistics and Philosophy on May 1st, 1991 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Abstract

In this thesis, I discuss the philosophical foundations of Hilbert's Consistency Programme of the 1920's, in the light of the incompleteness theorems of Gödel.

I begin by locating the Consistency Programme within Hilbert's broader foundational project. I show that Hilbert's main aim was to establish that classical mathematics, and in particular classical analysis, is a conservative extension of finitary mathematics. Accepting the standard identification of finitary mathematics with primitive recursive arithmetic, and classical analysis with second order arithmetic, I report upon some recent work which shows that Hilbert's aim can almost be realized.

I then discuss the philosophical significance of this startling fact. I describe Hilbert as seeking a middle way between two mathematically revisionary positions in the philosophy of mathematics - a kind of proto-intuitionism, and an extreme realism, associated with the views of Kronecker and Frege respectively. I outline a Hilbertian alternative to these positions. The result is a moderate realism that owes much to Quine. I defend it against certain objections, and display its virtues in a series of comparisons with alternatives currently influential in the literature.

In **Chapter Two**, I discuss the special status the Hilbertian gives to finitary mathematics. I argue that two ways of justifying this special status - by claiming that finitary mathematics is *ontologically* special, since it is committed only to expressions, and by claiming that finitary mathematics is *epistemologically* special, since its results are especially evident - are in fact hopeless. I then defend an alternative justification, drawing in part on Gödel's well known discussion of mathematical intuition.

In **Chapter Three**, I discuss the implications of incompleteness for the Hilbertian philosophy of mathematics. I argue, against some recent work by Michael Detlefsen, that the incompleteness theorems show definitively that Hilbert's Programme cannot be carried out in full generality. Drawing on recent work by Warren Goldfarb, I show that this conclusion follows from the First Incompleteness Theorem, and can be established without any controversial appeal to the semantic value of undecidable sentences. However, I argue that the fact of incompleteness adds to, rather than detracts from, the attractiveness of the basic Hilbertian position on the nature of mathematics.

Thesis Supervisor: Dr. George Boolos

Title: Professor of Philosophy

Acknowledgments

My interest in Hilbert was awakened by unpublished work by Michael Hallett, from which I have learned a great deal. I have since learned more from him in conversation. Hallett apart, my understanding of Hilbert (*qua* philosopher) owes most to the writings of Howard Stein and W. W. Tait.

In this thesis, I have some critical things to say about Michael Detlefsen's writings on Hilbert. That being so, let me emphasize here that I could not have written it without the stimulus provided by his valuable book on Hilbert's Programme. He also kindly showed me unpublished work, on the implications of the First Incompleteness Theorem, that helped me enormously.

Richard Cartwright, Joshua Cohen, and James Higginbotham all read and discussed parts of this manuscript with me, and provided detailed and helpful criticism. My supervisor, George Boolos, in addition to suggesting the project to me in the first place, was generous far beyond the call of duty with his time, his knowledge, his patience, and his fonts. I am grateful for all four, but I am especially grateful for the last two. Martin Davies has consistently been the most sympathetic, helpful, and stimulating interlocutor one could hope for. He read the manuscript, some parts of it many times, and provided copious and detailed criticism. He rarely tired of listening to me, and always seemed to understand my half baked ideas much better than I did. And not only did he do all this: he actually seemed to enjoy it.

I should mention a final debt, to an individual that is not a person. MIT provided me with the funds, and therefore with the opportunity to undertake a PhD in philosophy, at a time when the government of my own country, Great Britain, had decided that higher education in philosophy was a luxury we could no longer afford. I can only hope that, once that shameful decision is reversed, British universities will also dispense their research funds without keeping one eye ever on the nationality of deserving candidates.

CONTENTS

INTRODUCTION

This essay expounds a philosophy of mathematics closely akin to that associated with Hilbert's consistency programme of the 1920's. I try to explain the position, display its attractions, and defend it against some criticisms. I shall be particularly concerned to understand the *finitary standpoint* associated with Hilbert's Programme, in the light of what Gödel's discoveries reveal about the pervasive incompleteness of mathematical theories. **Chapter One** gives a critical discussion of the Hilbertian philosophy of mathematics. In **Chapter Two** I discuss the finitary standpoint, and in **Chapter Three** the significance of incompleteness.

Since the finitism I discuss in intended to be pretty much that advocated by Hilbert, I shall spend some time explaining and discussing Hilbert's views. But my primary purposes are not exegetical, and whilst I hope not to misrepresent Hilbert, I am more concerned that the philosophical position I try to articulate should seem plausible and attractive in its own right. Hilbert, it should be remembered, was not a philosopher, and although his prose writings are full of distinctively philosophical remarks, they contain only the barest outlines of a philosophy of mathematics. That outline can be fleshed out in many different ways, compatibly with all that Hilbert actually says. I find that much of the existing philosophical literature fleshes out the outline in ways which make Hilbert's position seem very weak. Since I don't find anything very weak in the outline itself, I thought it worthwhile to try to do better. For those interested in exegetical questions, **Appendix One** gives my views on the interpretation of Hilbert in more detail.

I have been asked more than once why I am so interested in a bankrupt programme. The answer, obviously, is that I do not think that the programme is bankrupt. To be sure, the incompleteness theorems show, definitively, that no interesting mathematical theory can be proved to be consistent by the finitary metamathematical techniques pioneered by Hilbert. In the sense that the specific goal of a finitary consistency proof for classical mathematics, and especially classical analysis, cannot be achieved, Hilbert's Programme failed. But what this means is that one specific goal, albeit a central one, must be given up. The consequences of this fact for the underlying view of mathematics are not at all clear cut, and this is one of the things I hope to show.

Furthermore, we now know that something very close to a finitary consistency proof can in fact be given for almost all of classical analysis. More precisely: recent work has demonstrated that almost all of classical analysis can be formalized in systems which are (finitarily) provably conservative over primitive recursive arithmetic. This constitutes a far reaching, if nevertheless partial, completion of Hilbert's Programme. I explain these recent discoveries more fully, and discuss their philosophical interest, in **Chapter One**.

CHAPTER ONE:
Hilbert's Programme and the Philosophy of Mathematics

Introduction: In this chapter, I discuss the philosophical foundations of Hilbert's Programme. In **Section One**, I sketch the Programme itself, giving a provisional and superficial account of the philosophical issues involved. **Section Two** then outlines some recent work in mathematical logic which shows that a significant *partial* realization of Hilbert's Programme is possible. (Full realization, of course, is precluded by the Gödel incompleteness theorems).

I then begin to examine the philosophical issues involved in more detail. In **Section Three**, which is partly historical in character, I describe the 'problematic' of Hilbert's Programme, whilst **Section Four** discusses Hilbert's own philosophical opinions. Then, in **Section Five**, I sketch a philosophical position which seems to me to comport with Hilbert's expressed views in all essentials, which is sensitive to the underlying concerns discussed in **Section Three**, and which is also, I hope, reasonably plausible and attractive in its own right. Finally, in **Section Six** I contrast this philosophical position with some currently influential alternatives.

Section One: Hilbert's Programme - A Preliminary Sketch. Standardly, Hilbert's Programme is described as a response to the crisis in the foundations of mathematics caused by the discovery of the paradoxes of naive set theory. Thus, even the most sympathetic of recent philosophical commentators on Hilbert's Programme, Michael Detlefsen, begins his book with the words: 'Hilbert's Program was founded on a concern for the phenomenon of paradox in mathematics'.[1] On this standard view, the discovery of the paradoxes of naive set theory had engendered a widespread suspicion that classical mathematics, and in particular classical analysis, might itself harbor similar paradoxes. Hilbert's Programme is then seen as an attempt to allay this sceptical doubt, by proving that the standard mathematical systems are consistent. Presented in this way, the primary objective of Hilbert's Programme is epistemological in character.

---

[1] See Detlefsen, Michael [1986], pix.

If I understand Hilbert aright, however, this standard view is misleading, and obscures many of the philosophically most interesting aspects of Hilbert's thought. Whilst the concern to combat doubts about the consistency of classical mathematics was there, I do not think that it was really fundamental. For to begin with, neither Hilbert, nor any other mathematician known to me, has *ever* given serious grounds for suspicion concerning the *consistency* of classical mathematics, although many mathematicians have found classical mathematics unsatisfactory in other ways.[2] And in particular, the intuitionistic attack on classical mathematics - the major stimulus behind the mature version of Hilbert's Programme - was not based on any suspicion that classical mathematics might be inconsistent. Brouwer in particular does not make this claim, and repeatedly stressed that a proof of consistency, in what he called 'Hilbert's formalist sense', would leave his objections to classical mathematics untouched.[3]

Steven Simpson gets much closer to the real roots of Hilbert's concerns when he writes that the aim of Hilbert's Programme was 'above all . . . *to clarify and justify the mathematician's use of the infinite* '.[4] The 'infinite' here, of course, is the 'completed' infinite of Cantorian set theory, as opposed to the 'potential' infinite accepted by all mathematicians.[5] Now, the 'phenomenon' of paradox in mathematics alluded to by Detlefsen is not unconnected to this Hilbertian project of clarifying and justifying the mathematician's use of the actual infinite. For the phenomenon in question is in fact the inconsistency of naive set theory, and set theory is the systematic study of the actual infinite. The discovery of the paradoxes of naive set theory, then, might well motivate a concern to clarify and justify the set theorists' use of the actual infinite. But Hilbert's desire to clarify and justify the mathematician's use of the actual infinite does not in fact originate in the set theoretic paradoxes. The fundamental concerns underlying the mature consistency programme of the 1920's already dominate his work in the 1890's on the foundations of geometry, a decade before the paradoxes were discovered.

---

[2] I know of at least two places in the literature in which the consistency of classical analysis is said to be suspect - in Weyl's semi-popular book 'Das Kontinuum', and in Genzen's essay 'The Present State of Research into the Foundations of Mathematics' (see Genzen [1969] p235). Neither author suggests that there is anything like scientific evidence that analysis might be inconsistent. The sole ground offered for suspicion is apparently some vague and unspecified resemblance between arguments in analysis and in naive set theory.

[3] See e.g. Brouwer [1923] p336.

[4] See Simpson, Steven [1988], p350.

[5] In Hilbert [1925], even analysis is said to involve only the 'potential' infinite. What Hilbert means here, of course, is that real numbers may be thought of as 'potentially' infinite *sets*. This is a more complex notion of potential infinity than that implicated in the simply infinite sequence of natural numbers.

Simpson's characterization of Hilbert's aims goes deeper than the common alternative, in that it suggests, quite rightly, that a reasonable assurance that some system of set theory is consistent - the kind of assurance we now have with respect to **ZF**, for example - would not suffice to allay Hilbert's foundational concerns.[6] To see why, we shall have to examine those concerns in detail. In particular, we shall have to get clear about the *use* to which Hilbert hoped to put a consistency proof for classical mathematics. For it is vital to understand that, in Hilbert's Programme, a consistency proof figures as a means to an end, rather than as an end in itself.

Hilbert proposed a two step programme, by means of which he hoped to show that the whole of classical mathematics could be reconciled to a foundationalist position - finitism - which nevertheless respected (what he believed to be) the legitimate aspects of the intuitionistic critique. In bold outline, the two steps are the following:

(**Step One**) Classical mathematics is to be *formalized* - set out as a deductive system (or series of deductive systems), with a clearly demarcated background logic and an effectively specified syntax.

(**Step Two**) The system or systems constructed in **Step One** are then to be provided with a *syntactic* consistency proof.

What we need to understand is the way in which this two step programme addresses the problem of clarifying and justifying the mathematician's use of the infinite.

The heart of the matter lies in **Step Two**, with its demand for a syntactic consistency proof.[7] In the ordinary way, one shows the consistency of a theory T by exhibiting a model of the axioms of T - an interpretation of the primitive vocabulary of T on which the

---

[6] **ZF** is the standard axiomatization of Zermelo/Frankel set theory, which includes the Axiom of Choice.

[7] The first part of this programme is apt nowadays to strike one as unproblematic. But this perspective was not shared by Hilbert's contemporaries, for whom an attempt at the formalization of all of classical mathematics would have seemed a rash and doubtful enterprize. Uncertain as to whether formalization would be possible at all, the early axiomatizers were understandably inclined to work in very powerful theories. The upshot of their efforts was the discovery that all of mathematics can be formalized in any of the standard set theories, **ZF** being the one most commonly used. Consequently, the first objective of Hilbert's Programme has been achieved. But **ZF** is a very powerful theory. Indeed, it has an enormous excess of power over anything strictly required for the execution of **Step One**. One of the interests of the research I shall describe in **Section Two** below, on partial realizations of Hilbert's Programme, lies in showing just how little set theory is involved in much of standard mathematics. The philosophical importance of this will emerge in **Section Four** and **Section Five**.

axioms of **T** are all true. This procedure yields a *semantic* consistency proof. Hilbert was well aware of the possibility of giving semantic consistency proofs for axiomatized theories. Indeed, he was one of the early pioneers of this technique, which features prominently in his early work on systems of arithmetic and geometry. Why, then, did Hilbert regard semantic consistency proofs inadequate for his foundational purposes? And are there any good reasons for refusing to rest content with semantic consistency proofs in metamathematics?

These are important questions, and they raise a number of large philosophical issues, which will be discussed at some length in **Section Three** through **Section Six** below (as well as throughout **Chapter Two**). For the moment, though, I will rest content with a partial answer, which has the merit of drawing our attention to the actual use to which Hilbert proposed to put a syntactic proof of consistency.

A syntactic consistency proof takes the form of an inductive argument (typically on the length of proofs) demonstrating that some chosen absurdity (e.g. $\ulcorner 0=1 \urcorner$) is not a theorem of the system. Since $\ulcorner 0=1 \urcorner$ *is* derivable in any *inconsistent* system, such a demonstration amounts to a proof of consistency. Now, Hilbert divided classical mathematics into two parts, a real (or *finitary*) part, the consistency of which he regarded as unproblematic, and an *ideal* part (or parts) for which a consistency proof was required. Tentatively, let us identify the finitary part of mathematics with Primitive Recursive Arithmetic (PRA), and the ideal part with everything else. The interesting twist in the Consistency Programme then lies in Hilbert's demand that the consistency of the ideal part(s) be proved *using the resources of PRA only*. Notice that this is not, prima facie, an unreasonable demand. Proofs - finite, discrete, effectively constructed arrays of symbols - are themselves mathematical objects of exactly the kind dealt with in PRA. Thus whilst the content of some theory in the ideal part of mathematics may very well transcend the bounds of finitary mathematics, the proof structure of the theory will not. By studying the proof structure of an ideal theory directly, and independently of any model for the theory, Hilbert hoped to provide an assurance of consistency *which did not involve any appeal to the theory's (infinitary) mathematical content.* The focus on a syntactic consistency proof, then, was intended to dispense with any appeal to models in completed infinities.

Assume, then, that we have partitioned classical mathematics into real and ideal parts. The technical result sought by Hilbert's Programme can now be described as *a proof that the use of ideal mathematics does not enable us to prove any new real formulas* - any real

formulas, that is, that do not have proofs which contain only real formulas. Obviously, there will also be proofs of theorems containing ideal elements, and it was no part of Hilbert's aim to show that the ideal elements could be eliminated from *these* proofs. Hilbert is not a reductionist in the naive sense. The justification for the use of ideal elements, though, was to be provided by a demonstration that they could not prove any real formulas that did not have alternative real proofs. This is Hilbert's *Conservation Programme*, so-called since its goal is to prove that the full system of classical mathematics, real elements included, is a conservative extension of the finitary fragment. The twist, remember, is that the conservation property is to be proved for the whole system by a *real proof*: conservation must be shown to be *finitistically* provable.

The link between the Conservation Programme and the two part Consistency Programme is then forged by an argument I shall call the *Master Argument*. As we shall see, this argument purports to show that, if we have a finitary consistency proof for an ideal system **I**, then for any finitary formula S provable in **I**, we can effectively construct *a purely finitary proof of S*. Think for a moment of what this would mean, if it were true. Much of our deepest knowledge of the natural number series - for example, most of what we know about the distribution of prime numbers - comes from the subject known as analytic number theory, in which the ideal methods of real analysis are brought to bear on purely number-theoretic questions. Sometimes, but not often, these analytic methods can be shown to be eliminable.[8] That is to say, it can sometimes be shown that the use of analytic ('ideal') methods is not essential to establishing some elementary ('finitary') result. When this is so, the ideal elements used in the analytic proof may be thought of as merely simplificatory devices, abbreviating an enquiry that could, in principle, be conducted in purely finitary terms. But this is not the typical case. Most often, the use of ideal elements, so far as we know, is ineliminable. A successful prosecution of the Master Argument would therefore be of very great mathematical significance, since it would not only show, *with full generality*, that the use of ideal elements was purely simplificatory in character, it would also provide mathematicians with a powerful technique for obtaining new proofs. Since it is often (although by no means always) the case that finitary proofs provide more information than ideal ones, the Master Argument wears the appearance of being a research tool rich in mathematical potential. Even a constructivist mathematician - Brouwer, as it might be - could not object to research devoted to discovering the

[8] The paradigm here is Kronecker's finitary version of Dirichlet's analytic proof that any arithmetic series containing two relatively prime terms contains infinitely many primes. I shall have more to say about this in **Section Three** below.

mathematical properties of infinitary, ideal elements, since such research would have been shown to have the potential of providing nothing but constructively acceptable results, if applied in the part of mathematics the constructivists care about.

Fix now some ideal system **I**, and some finitary subtheory **F** of **I**. Assume that the expressions of **F** and **I** have been coded in some reasonable way. I use the expression '$\ulcorner \varphi \urcorner$' to denote the code number assigned to the expression 'φ'.[9]

I am now in a position to state the master argument. By the construction of Goodstein's logic free' formalization of PRA described in **Appendix Two**, it is clear that **PRA** may be thought of as essentially a free variable calculus.[10] (Closing a formula by prefixing universal quantifiers leaves its quantificational complexity unchanged, and existential quantifiers may be thought of as having been replaced by primitive recursive functions in Skolem's way.) The finitary formulas, then, are equations between primitive recursive terms. Let S be some finitary formula $f(x^{\rightarrow}) = g(x^{\rightarrow})$. and suppose

(1) $\mathbf{I} \vdash \ulcorner S \urcorner$

Suppose now that we have a finitary consistency proof for **I**. Then we will have

(2) $\mathbf{F} \vdash \neg(\mathrm{Bew}_{\mathbf{I}}(x^{\rightarrow}, \ulcorner z \urcorner) \wedge \mathrm{Bew}_{\mathbf{I}}(y^{\rightarrow}, \underline{\neg}(\ulcorner z \urcorner))$

for arbitrary $x^{\rightarrow}$, $y^{\rightarrow}$, where '$\mathrm{Bew}_{\mathbf{I}}$' is a primitive recursive predicate which numeralwise expresses the provability relation of **I**, and '$\underline{\neg}(\ulcorner z \urcorner)$' denotes the value of the arithmetical correlate of the negation operation for the value $\ulcorner z \urcorner$. By the properties of encoding in **F**, we have

(3) $\mathbf{F} \vdash \underline{\neg}\ulcorner G \urcorner = \ulcorner \neg G \urcorner$

for any G, so we identify the numerical value of performing the arithmetic operation corresponding to logical negation on the formula G with the code number of the negation of

---

[9] I discuss syntactic conventions at greater length in **Appendix Two.**

[10] **PRA** is the standard formalization of PRA, given in Robbin [1967] e.g., and reproduced in **Appendix Two.** In the same way, **ZF** is the standard formalization of ZF. **PA**, on the other hand, is the standard formalization of arithmetic.

G. In virtue of the properties of **PRA**, for any given terms f and g, we can effectively construct a term h such that

(4) $\mathbf{F} \vdash f(x^{\rightarrow}) \neq g(x^{\rightarrow}) \rightarrow \mathrm{Bew}_{\mathbf{I}}[h(x^{\rightarrow}), \dot{\neg}(\ulcorner f(x^{\rightarrow}) = g(x^{\rightarrow}) \urcorner)]$.

Then by (1), for some k, k is the code of a proof in **I** of S, whence

(5) $\mathbf{F} \vdash \mathrm{Bew}_{\mathbf{I}}(\mathbf{k}, \ulcorner S \urcorner)$

i.e.

(6) $\mathbf{F} \vdash \mathrm{Bew}_{\mathbf{I}}(\mathbf{k}, \ulcorner f(x^{\rightarrow}) = g(x^{\rightarrow}) \urcorner)$.

But then from the consistency assumption (2),

(7) $\mathbf{F} \vdash \neg[\mathrm{Bew}_{\mathbf{I}}(h(x^{\rightarrow}), \dot{\neg}(\ulcorner f(x^{\rightarrow}) = g(x^{\rightarrow}) \urcorner)) \wedge \mathrm{Bew}_{\mathbf{I}}(\mathbf{k}, \ulcorner f(x^{\rightarrow}) = g(x^{\rightarrow}) \urcorner)]$

whence from (6),

(8) $\mathbf{F} \vdash \neg \mathrm{Bew}_{\mathbf{I}}(h(x^{\rightarrow}), \dot{\neg}(\ulcorner f(x^{\rightarrow}) = g(x^{\rightarrow}) \urcorner))$

Therefore, from (4) and (8) by modus tollens,

(9) $\mathbf{F} \vdash f(x^{\rightarrow}) = g(x^{\rightarrow})$

i.e., $\mathbf{F} \vdash \ulcorner S \urcorner$.

As promised, a finitary consistency proof of **I** yields a finitary proof of any finitary formula provable in **I**. Given the Master Argument, then, Hilbert has a fully cogent response to Brouwer's criticisms of the infinitary parts of classical mathematics.

Section Two: Partial Realizations of Hilbert's Programme: Even in this brief and approximate sketch, it is apparent that Hilbert's proposal for the clarification and justification of the use of the actual infinite is subtle, sophisticated and, initially at least, by no means implausible. Still, these attractions are likely to seem insignificant when set alongside one large and uncomfortable fact. The Master Argument requires finitary

consistency proofs for ideal theories, and there are none. This raises the question which is immediately asked whenever one makes claims about the philosophical importance of Hilbert's Programme - the question: What philosophical interest can this defunct programme possibly have? Given that Gödel has shown that one cannot prove consistency finitistically for any interesting mathematical theory, surely Hilbert's Programme can be of historical interest at best?

I do not think so. For it is now known that, despite incompleteness, the Conservation Programme loosely described above can in fact be carried out for almost all of classical analysis.[11] This is a very remarkable discovery in the foundations of mathematics, and I shall devote this section to explaining it in (slightly) greater detail. To do so, we shall have to describe the Hilbertian project into a more mathematically precise way.

We shall continue to identify the finitary part of classical mathematics with PRA. Now, once real and ideal parts of mathematics have been distinguished in this fashion, it might seem that the task that Hilbert faces is that of proving that all of contemporary mathematics is conservative over **PRA**. But I do not think that that is fair to Hilbert, since mathematics as we now have it differs very radically in extent, and indeed in character, from mathematics as Hilbert knew it. It seems to me much more reasonable to see Hilbert as attempting to clarify and justify the use of the actual infinite in what I shall call *ordinary mathematics* - to a first approximation, the mathematics done by mathematicians who are not set theorists.[12] More precisely, ordinary mathematics excludes those parts of mathematics that rely very heavily on the abstract theory of ordinal and cardinal number. In Simpson's classification, for example, ordinary mathematics includes geometry, number theory, calculus, differential equations, real and complex analysis, countable combinatorics, and some parts of topology.[13] Ordinary mathematics does not include infinitary combinatorics, general topology, uncountable algebra. This distinction is, of

---

11 Sieg, W. [1988] gives a useful survey of the relevant work.

12 The idea that there are mathematicians who are not set theorists is likely to strike some philosophers as very odd, since philosophy has been much impressed by the fact that pretty much all of mathematics can be modelled in set theory. This fact has encouraged the philosophical conviction that mathematics just *is* set theory, albeit set theory in disguise. And if mathematics is set theory, what can it mean to say that there are mathematicians who are not set theorists? But this is just a confusion. It is one thing to show that the rings, modules, ideals etc. discussed by algebraists can be shown to be sets of a certain kind. It is quite another thing to show that the algebraic properties of those objects are best studied by the theory of sets. The confusion here is analogous to that made in natural science by very strong forms of physicalism: the fact that the objects studied in biology are physical objects has no tendency to show that the biological properties of those objects are best studied by physicists.

13 See Simpson [1988], p432.

course, vague: descriptive set theory is an important border line case. But vague as the distinction is, mathematicians in practice appear to understand it reasonably well. What is more, the distinction permits the confident identification of a particular formal system as the ideal mathematics with respect to which Hilbert aimed to justify and clarify the use of the actual infinite. That system is $\mathbf{Z_2}$, the system of second order arithmetic described in **Appendix Two.**

We are now in a position to give a much more mathematically precise characterization of Hilbert's Programme. The Programme has the following three parts:

(HP1) Formalize finitary mathematics - i.e. formalize primitive recursive arithmetic.

(HP2) Formalize ideal mathematics - i.e. formalize second order arithmetic.

(HP3) Give a finitarily correct consistency proof for second order arithmetic.

With **(HP1)**-**(HP3)** in hand, the Master Argument would then demonstrate the realizability of the Conservation Programme by proving **(Conserv)**:

(Conserv) $\mathbf{Z_2}$ is conservative over **PRA** with respect to Π_1 formulae (i.e. w.r.t. finitary formulae).

A proof of **(Conserv)** is the fundamental goal of Hilbert's Programme. As I shall now show, something quite remarkably close to it has in fact been achieved.

In the 1970's, Harvey Friedman and his associates investigated a subsystem of $\mathbf{Z_2}$ known as $\mathbf{WKL_0}$. $\mathbf{WKL_0}$ is a theory in the language of $\mathbf{Z_2}$, but it draws upon only a very limited part of the logical resources of full second-order logic. In particular, abstraction and induction are allowed *only for Σ_1 formulas* (with respect to which **R**, and hence **PRA** are complete).[14] $\mathbf{WKL_0}$ includes **PRA**, but also includes an infinitary axiom known as *weak König's Lemma*.[15] As a consequence, several important non-constructive theorems of classical mathematics are provable in $\mathbf{WKL_0}$, including the theorems that establish the basic properties of continuous functions of several real variables, and the local existence

---

[14] See **Appendix Two** for references.

[15] Weak König's Lemma asserts that any infinite tree of finite sequences of zeros and ones has an infinite path.

theorem for solutions of systems of differential equations. In spite of this, $\mathbf{WKL_0}$ is conservative over **PRA** with respect to Π_1 formulae, and this conservation property can be proved in **PRA**.[16] (Indeed, $\mathbf{WKL_0}$ is conservative over **PRA** with respect to Π_2 formulae.)

More recently, Simpson and his associates have investigated a system $\mathbf{WKL_0^+}$ which adds to $\mathbf{WKL_0}$ an additional, stronger infinitary axiom.[17] It has been shown that $\mathbf{WKL_0^+}$ is also conservative over **PRA** with respect to Π_2 formulae, and this conservation property is once more provable in **PRA**. The extra infinitary resources of $\mathbf{WKL_0^+}$ make it possible to prove several additional nonconstructive theorems of functional analysis which appear to be unprovable in $\mathbf{WKL_0}$. In his June 1988 report on this ongoing research, Simpson suggests that it may be possible to define still stronger subsystems of $\mathbf{Z_2}$ which will prove still further theorems of infinitistic analysis, whilst remaining provably conservative with respect to **PRA**. We have here, therefore, a far reaching and as yet incomplete partial realization of Hilbert's Programme.[18] Whilst (**Conserv**) has not been, and cannot be, established in full generality (since that would imply the finitarily provable consistency of **PRA**, contra the Second Incompleteness Theorem), it can be, and has been, established for much of the mathematics that Hilbert most wanted to protect against the intuitionist challenge.

In virtue of these results, notice, it appears likely that at least a very great part of the mathematics used in natural science can be formalized in conservative extensions of **PRA**. I do not know if all of the mathematics used in natural science can be formalized in this way, and I suspect that this question may not have a determinate answer. At the outer limits of theoretical physics, my grip on what counts as natural science and what counts as mathematics goes hazy. Still, it is clear that the great bulk of applicable mathematics has the character Hilbert hoped to establish for all of ordinary mathematics. To my mind, this alone justifies the claim that Hilbert's Programme is of more than historical interest.

---

[16] See Friedman [1976].

[17] Let $2^{<N}$ denote the set of all finite sequences of zeros and ones. Then Simpson's axiom says that, given any sequence of dense subsets of $2^{<N}$ which is arithmetically definable from some given set, there exits an infinite sequence of zeros and ones which meets each of the given dense subsets. For more details on the theorems provable in the resulting system, see Simpson [1988]. The mathematical details have not yet appeared in print, to the best of my knowledge. They will be published in Simpson's 'Subsytems of second order arithmetic' (forthcoming).

[18] What is more, I think it of considerable philosophical interest to observe that much of the mathematics basic to physical science seems to be formalizable in these finitarily conservative systems. I shall return to this point in **Section Four** and **Section Five** below.

Section Three: The Context of Hilbert's Programme. Let us recapitulate. We have characterized Hilbert's project as that of clarifying and justifying the mathematician's use of the actual infinite. We have seen that the clarification and justification was to take the form of a demonstration that ideal mathematics was a conservative extension of a finitary mathematics in which the only infinity was the potential infinity of the natural number sequence. We have elaborated the quasi-technical argument that is the backbone of Hilbert's proposal as to how this job of clarification and justification is to be done. And we have seen that the use of the actual infinite in a surprisingly extensive part of ordinary mathematics can be shown to satisfy Hilbert's conservation requirement over a part of mathematics that makes no use of completed infinities.

What we do not yet have, however, is any real sense of what the philosophical interest of all this is supposed to be. What exactly is the problem with completed infinities to which Hilbert's ingenious proposal is supposed to be a solution? Why should philosophers of mathematics care which parts of mathematics are and are not conservative over **PRA**? Why, for that matter, should mathematicians? To which interesting questions in the philosophy of mathematics is all this algebra relevant, and why? This section begins to address these questions, by describing what Hilbert took to be the philosophical problems to which his programme was intended as a response. What I hope to do now is uncover the (fragmentary, but highly suggestive) philosophy of mathematics that underlies Hilbert's Programme. I take as my point of departure the historical context of Hilbert's foundational work.

At the time of his first engagement with foundational issues - the period from around 1898 to 1904 - we find Hilbert fighting in a battle on two fronts.[19] On one front, he is engaged with the constructivist Kronecker, whilst on the other, the foe is the arch realist Frege. In both cases, Hilbert was faced with a *revisionist* philosophy of mathematics - one which implied (indeed, which explicitly argued) that much of the recent mathematical research most valued by Hilbert was to be rejected as without genuine scientific value. In both cases, the locus of the controversy was geometry. On the front attacked by Kronecker, *all*

---

[19] 1898, since in that year Hilbert gave his first public lectures on Euclidean geometry - the lectures which formed the basis of *Die Grundlagen der Geometrie*, published two years later. According to Reid, however, the foundational issues he engages in those lectures, together with the basic position he adopts towards them, began to occupy his mind several years earlier. I give 1904 as the terminus of this early engagement with foundations since that year marks the publication of the article 'Uber die Grundlagen der Logic und der Arithmetik', after which Hilbert published nothing explicitly on foundations until 1917.

systems of geometry were under attack, for Kronecker's foundational position implied that geometry as such was not part of pure mathematics.[20] Frege was only slightly less radical, allowing Euclidean geometry, but only Euclidean geometry, as part of pure mathematics. If either view was accepted, all of non-Euclidean geometry, including the profound discoveries that had transformed the entire mathematical landscape of the later nineteenth century, would have to be given up.

The initial impetus for the controversy was provided by the realization, widespread by the mid-nineteenth century, that there existed many consistent alternatives to Euclidean geometry. Coupled with the conviction that geometry was, by definition, that part of mathematics dealing with a particular kind of mathematical objects - objects such as points, lines, planes etc. - the existence of consistent alternatives to Euclid took on a puzzling aspect.

Consider the famous puzzle associated with the Euclidean parallels postulate, for example. It is an axiom of Euclidean geometry that, given a line L and a point P not incident with L, there exists exactly one line L' incident with P coplanar with L. Since Euclid's own time at least, the status of this axiom had been controversial. It was widely held to lack the immediate obviousness of the other Euclidean axioms, and this prompted many generations of mathematicians to attempt to demonstrate that it is in fact a consequence of the other four Euclidean axioms (rather than an axiom in its own right). The mathematicians instincts here were quite correct, notice - there is something less than obvious about the parallels axiom. Euclidean geometry, as we now know, leaves the crucial primitive notion of *congruence* insufficiently determinate. Consistent alternatives to Euclid can be generated by leaving the remaining Euclidean axioms unchanged, and introducing conditions which have the effect of rendering congruence a more determinate notion. But this idea of allowing alternative interpretations of the primitive expressions of a theory, natural as it is to us now, was profoundly alien to the prevailing conception of mathematics in the early nineteenth century. For on that prevailing conception, the primitives of Euclidean geometry had fixed senses, understandable independently of geometric theory, in virtue of which the Euclidean axioms were true - where this does not mean, true on the intended interpretation of Euclidean geometry, but just, TRUE.

---

[20] This is a little too strong, since the finite geometries with which we are now familiar would escape Kronecker's strictures. As we shall see in more detail below, Kronecker's position led to the rejection of geometry insofar as geometry depended upon the Weierstrass/Dedekind notion of continuity.

However, by the mid-nineteenth century, it was known that consistent alternatives to Euclid (systems, that is, that satisfy the Euclidean axioms without the parallels postulate) allow the existence of either no such line L', or many such lines: the parallels postulate is independent of the other axioms, and hence cannot be derived from them. Given the prevailing conception of geometry as that part of mathematics which deals with fixed, univocal notions of point, line, plane, incidence etc., this situation seemed very puzzling. For on that conception, at most one system could be telling us *the truth* about parallel lines in the plane. But given the consistency of alternatives to Euclid, there appeared to be no genuinely mathematical way of telling which.

One reaction to this multiplicity of consistent geometries implied the exclusion of geometry from pure mathematics. This was the position Hilbert associated with Kronecker, but it was all the more influential for having the authority of no less a figure than Gauss. In a well known letter written in 1830, Gauss states

> According to my deepest conviction, the theory of space has a quite different position in our apriori science from that of the pure theory of magnitudes. Our knowledge thoroughly lacks that complete conviction of its necessity (and thus of its absolute truth) which is the characteristic of the latter. We must in all humility admit that *if number is merely the product of our intellect, space on the other hand has a reality outside of this*, a reality to which we cannot completely prescribe laws in an apriori way. (Gauss to Bessell, 30 April 1830, my emphasis.)[21]

This drastic resolution of the difficulty posed by alternative geometries must not be confused with the more familiar alternative of counting *all* the consistent geometries as acceptable parts of pure mathematics, but involving shifting interpretations of the geometric primitives. The crucial difference, of course, is that the Gauss/Kronecker alternative clings to the traditional understanding of pure mathematical theories as theories with a *fixed interpretation.* Geometry has been relegated from this status in virtue of the fact that the unique *correct* interpretation of the primitives of geometry cannot be given by mathematical means. The Gaussian demotion of geometry is based upon the conviction that only physical experimentation (or at any rate, some aposteriori element) can determine this correct interpretation of the geometric primitives. Correctness being beyond the reach of mathematics as such, geometry becomes relegated to the status of mechanics - part of theoretical physics rather than pure mathematics.

---

[21] See Gauss [1880] p497.

The problem here, evidently, stems from the acceptance of a certain foundational thesis that almost no-one now believes, viz.

(K1) Mathematical theories must always be theories *with a fixed interpretation.*

Indeed, one reason why almost no-one now believes **(K1)** is that we have almost all learned the lessons taught by the realization that there are consistent alternatives to Euclid.[22] But independently of how one feels about **(K1)**, it will surely strike the reader that **(K1)** alone cannot suffice to drive all of geometry from the realm of pure mathematics. For as every high school student of analytic geometry in effect knows, geometric systems such as Euclid's can be modelled in the theories of the real and complex numbers. So even if our geometric intuition has proved inadequate to the task of securing a purely mathematical content for geometric theories, we can save geometry for pure mathematics by appealing to analytic interpretations of geometric systems.

It is at this point, though, that Kronecker's position becomes both very interesting, and very important for the attempt to understand Hilbert. For Kronecker's position on geometry was determined, not by **(K1)** alone, but by the following extension of **(K1)**

(K2) Mathematical theories must always be theories with a fixed interpretation, *and that interpretation must be given as a decidable construction from the natural numbers.*[23]

---

[22] I say almost all, since as recently as 1980 an article appeared in the philosophical literature arguing that the standard proofs of the independence of the parallels postulate do not answer the traditional question which provoked the long history of attempts to derive the parallels postulate from the remaining axioms of Euclidean geometry. Geoffrey Hunter [1980] seems to think that there really is a further question to be answered, which (if I understand him correctly) is something like the question whether the Fregean Thought expressed by the parallels postulate really is TRUE. The lesson taught by this episode in the history of mathematics, it seems to me, is that there is simply no answer to this question, since there is simply no such Thought.

[23] **(K2)** brought Kronecker and the young Hilbert into direct conflict, since (as we shall see) **(K2)** proscribes pure existence proofs, and Hilbert's first major result, the proof that a 'finite number in (rational) integers of invariants, by which all the rest of such invariants can be integrally represented' [Hilbert 1917 p196] always exists, was a pure existence proof. Hilbert's proof astonished the international mathematical community with its (comparative!) simplicity and brevity. Earlier attacks on subcases of the general problem solved by Hilbert had proceded by actually calculating the required basis, via algorithmic procedures of enormous complexity. Hilbert's proof, on the other hand, avoids such procedures entirely. The existence of a finite basis is proved indirectly, by a demonstration that a contradiction results from the assumption that no finite basis exists. Consequently, the proof gives no information as to how a basis might actually be found - it is completely non-constructive. To many of the leading mathematicians in this field, including Kronecker, Hilbert's proof was, in the mathematician Lindemann's famous phrase, 'unheimlich' - uncanny, wierd - and this in spite of the fact that, as we have already seen, results of this kind had been established fifty years earlier by Dirichlet. Most mathematicians quickly came to see the value of Hilbert's work on invariants, but Kronecker, in virtue of his acceptance of **(K2)**, never did. For him, work of this kind was simply not mathematics.

It is important to notice that (**K2**) constrains the *introduction of objects/concepts* in mathematics. Thus, if the concept of the *limit of a convergent series* is to be introduced, for example, (**K2**) makes any definition which does not enable one to determine whether or not a given number is the limit of a given convergent series illegitimate. (**K2**) also constrains attempts to prove that an object meeting a given specification exists, for according to (**K2**), any such proof must give us an *effective procedure* for finding an object meeting the specification.

Kronecker appears to have believed that a good deal of mathematics as he knew it could be recast in accordance with (**K2**). In one of the few passages in his published work which is not purely mathematical in content, he writes

> . . . one day it will be possible to 'arithmetize' the whole content of all the mathematical disciplines, i.e. to found them on the number concept alone with this taken in the narrowest sense, thus to cast off again the modifications and extensions which this concept has undergone,* *modifications and extensions which have been occasioned by the applications to geometry and mechanics.*
>
> * I mean here namely the addition of irrational and continuous magnitudes.[24]

Indeed, Kronecker took himself to be engaged upon exactly this radical programme of reconstruction (what we might call Kronecker's Programme). However, he himself explicitly asserted that the concept of *limit* resists this kind of treatment - in his phrase, 'remains alien to number theory'. What this in fact means, though, is that the limit *is not reducible to the finitary theory of the natural numbers.* Given that this is indeed so, the evident implication of adherence to Kronecker's strictures, so far as the mathematical community as a whole was concerned, was the loss of vast areas of well established and valuable mathematics (along with all of Cantor's nascent set theory). This devastation was undoubtedly the consequence of Kronecker's position as Hilbert understood it, and the prime reason why he opposed it so strongly.

'Decidable' and 'construction' now have precise technical senses, which were of course unknown to Kronecker. Still, he seems to have had in mind pretty much what we would mean by decidability and constructiveness. For Kronecker, if a definition is to be acceptable, it must be possible for us to determine in a finite number of steps whether an arbitrary object (of the right kind, perhaps) satisfies the definition. Thus, for example, he rejected Weierstrass's definition of 'irrational number' on the grounds that the definition did not in general enable one to determine whether an arbitrary sequence defined an irrational number.

[24] Kronecker [1887] p253, my emphasis.

It is not difficult to see how adherence to (**K2**) compelled Kronecker to exclude geometry from pure mathematics. Kronecker knew of several well-established mathematical results which apparently violate (**K2**), including, crucially, the central results of analytic number theory. This field had its inception in 1837, when Dirichlet (Kroneker's teacher, whom Kronecker revered) proved the theorem mentioned in Section Two above - that there are infinitely many prime numbers in any arithmetic progression which contains two relatively prime terms. This argument, as Dirichlet presents it, makes free use of continuous variables and limits: consequently, it violates (**K2**).

The importance of this theorem, and the field which it opened up, was noted by Dirichlet himself, when he wrote '[the] method I employ seems to me above all to merit attention by *the connection it establishes between the infinitesimal Analysis and the higher Arithmetic* ' (my emphasis). That there might be such a connection at all was in itself thought to be remarkable. Since antiquity, it had been believed that there were two quite different kinds of 'quantity' studied in mathematics - the continuous and the discrete. In analytic number theory, however, this distinction seemed to become blurred.

Dirichlet himself, though, did not think that his work had this effect. His own views on the connection are stated very clearly by another of his pupils, and Kronecker's arch rival, Richard Dedekind. In the Preface to the First Edition of his great essay *Was sind und was sollen die Zahlen?* Dedekind writes:

> From this point of view [the point of view adumbrated in *Was sind und was sollen die Zahlen*, that is] it appears as something self-evident and not new that *every theorem of algebra and higher analysis, no matter how remote, can be expressed as a theorem about natural numbers* - a declaration I have heard repeatedly from the lips of Dirichlet.[25]

This might reasonably be taken to mean that those parts of analytic number theory that appeared to make use of the theory of continuity essentially do not in fact do so - that strictly arithmetical proofs of all such theorems must exist.[26] This was undoubtedly the position attributed to Dirichlet by Kronecker, who took particular pride in his recasting, in 1885, of the proof of Dirichlet's theorem in conformity with (**K2**) - perhaps his major

---

25 Dedekind [1887] p25, my emphasis.

26 Of course, we now know that what Dedekind had partially discovered was not the reducibility of analysis to number theory, but the reducibility (in a different sense) of both to set theory. In reading Dedekind, Kronecker and indeed the early Hilbert, it is important to remember that they were unaware of the fundamental difference between reduction to a theory which permits quantification over numbers alone on one hand, and theories which permit quantification over sets of numbers on the other.

contribution to the project mentioned above as Kronecker's Programme. Notice that this suggests that Kronecker's position on (K2) was in fact flexible. He seems to have been prepared to allow that at least some non-constructive *arguments*, such as this one of Dirichlet, might have mathematical value - as stop-gaps on the road to a constructive proof, if nothing else. He was unbending, though, on the use of non-constructive *definitions*.

Why was it thought so important, though, that the use of limits, continuous functions etc. be shown to be eliminable from arguments in arithmetic? Once again, the answer that would have undoubtedly been given by Kronecker to this question in fact appears in print in the work of his arch rival, Dedekind. At the beginning of his monograph *Continuity and Irrational Numbers*, Dedekind writes of his dissatisfaction as a teacher at the 'lack of a really scientific foundation for arithmetic'. He continues:

> In discussing the notion of the approach of a variable magnitude to a fixed limiting value, and especially in proving the theorem that every magnitude which grows continually, but not beyond all limits, must certainly approach a limiting value, I had recourse to geometric evidences. Even now such resort to geometric intuition in a first presentation of the differential calculus, I regard as exceedingly useful, from the didactic standpoint . . . But that this form of introduction into the differential calculus can make no claim to being scientific, no one will deny.[27]

But what is so bad about the resort to the 'geometrically evident' in this context? Dedekind's answer to this is very revealing. In Section Three, he gives what he calls the 'essence of continuity' in the following 'axiom':

> If all points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes, this severing of the straight line into two portions.[28]

And then there comes the crucial passage:

> . . . I think I shall not err in assuming that every one will at once grant the truth of this statement; the majority of my readers will be very much disappointed in learning that by this commonplace remark the secret of continuity is to be revealed. To this I may say that I am glad if every one finds the above principle so obvious and so in harmony with his own ideas of a line; for I am utterly unable to adduce any proof of its correctness, nor has anyone the power. The assumption of this property of the line is nothing else than *an axiom by which we attribute to the line its continuity*, by which we find

---

[27] Dedekind [1872] p1.
[28] Dedekind [op cit] p11.

> continuity in the line. *If space has at all a real existence it is not necessary for it to be continuous*; many of its properties would remain the same even were it discontinuous.[29]

Later, in *Was sind und was sollen die Zahlen?*, Dedekind outlines a detailed mathematical defense of this last claim, by showing (as we would put it) the existence of a model of Euclid's axioms in which all ratios of lengths of line segments are *algebraic numbers*. Such a space is not, of course, continuous (the class of algebraic numbers is countable). Insofar as Euclid captures our 'geometric intuition', then, *our geometric intuition is not adequate to the task of capturing what is distinctive of continuity*. Even allowing for the existence of something like geometric intuition as traditionally conceived, nothing in our geometric intuition can suffice to provide the grounds for a scientific notion of continuity, since we do not, and perhaps cannot, experience space directly as continuous.

This does not mean that our conception of continuity, as Dedekind construes it, has *no* connection with what is geometrically evident, notice. What it means is that the introduction of the key concept in any developed mathematical theory of continuity - the concept of limit - cannot be given solely in terms of geometric notions available independently of the developed theory of continuity itself. Once matured, the theory has no point by point contact with the geometrical notions in which it originates. In Dedekind's words, we *attribute continuity to the line* by the imposition of this axiom.

But whilst Kronecker was in agreement with Dedekind with respect to the inadequacy, from a mathematical point of view, of our geometric intuition, he could not accept the attempts to rescue geometry for mathematics via the kind of theory of continuity that Dedekind outlined in this passage. For that theory, elaborated by Weierstrass, Dedekind and Cantor in succession, provided for Kronecker the paradigm case of a completely unacceptable violation of **(K2)**.[30]

What we have now introduced are the horns of a trilemma - a trilemma which Kronecker, unable to contemplate abandoning **(K2)**, took to force the exclusion of geometry from pure mathematics. Geometric theories, to count as mathematics at all, must have a fixed, stable

---

[29] Dedekind [op cit] pp11-12, emphases mine.

[30] The violence of Kronecker's opposition to Cantor's methods in particular is legendary - and it appears, unfortunately, that it must remain so; for so far as I can discover, Kronecker nowhere denounces Cantor in print. With Dedekind, however, the situation is different. In a paper of 1886, on Dedekind's algebraic number theory, Kronecker is direct and uncompromising in his rejection, and it is quite explicit that he rejects Dedekind's work on the grounds that it violates (K2).

content - a unique and specifiable subject matter. But what was this subject matter? We might, on the one hand, give 'line', 'point', 'plane', 'incidence' a physical interpretation; but then geometry is not pure mathematics, since the properties of the geometrical objects, thus interpreted, can only be empirically determined - this is sufficiently demonstrated by the existence of consistent alternatives to Euclid. On the other hand, we might appeal to a priori geometric intuition. But intuition had proved to be inadequate to the task of fixing the sense of the geometric primitives, and necessarily so, since Dedekind had shown that the crucial concept of continuity could not be grounded in intuition alone. Finally, we might appeal to the Weierstrass/Dedkind/Cantor theory, and thus recover for geometry the essential continuity concepts; but these theories are not acceptable as mathematics, in virtue of violations of (**K2**). These three possibilities seemed exhaustive, since Kronecker, for all his conviction that genuine analysis could always be 'arithmetised', had no suggestions to make with respect to an alternative account of limits in accordance with (**K2**). Consequently, there was no room for geometry in pure mathematics.

This consequence, one might think, ought to give anyone grounds for abandoning adherence to (**K2**), rather than abandoning geometry. However, we should try to understand why Kronecker (and many other mathematicians after him) took the other option, because in so doing, we stand to learn something very important about Hilbert's attitude to Kronecker's Programme, and thus about Hilbert's understanding of what he calls finitary mathematics. Unfortunately, Kronecker's motivations have to be gathered almost entirely at second and third hand, from reports of his views from friends and colleagues - it is from such reports that we have Kronecker's famous dictum: 'The good Lord made the natural numbers; all the rest is the work of man'. But even if the attribution of this *bon mot* to Kronecker is correct, we must still be puzzled about what it means. Why should the divine origin of the natural numbers dictate acceptance of (**K2**), at the cost of abandoning so much mathematics?

In one place in his writings, however, Kronecker gives us something more solid to go on. In his essay *Über den Zahlbegriff*, he makes the following very revealing remark on the epistemology of mathematics, once more invoking the authority of Gauss:

> The difference in principles between geometry and mechanics on the one hand, and the remaining mathematical disciplines, here comprised under the designation 'arithmetic', consists according to Gauss in this, that the object of the latter, Number, is solely the product of our mind, whereas Space as well as Time have

> also a *reality, outside* our mind, whose laws we are unable to prescribe completely a priori.[31]

This is of course the very claim that inclined Gauss to banish geometry from pure mathematics: the real content of geometry (and mechanics) cannot be known a priori, where we may take that to mean, cannot be determined by mathematical investigation alone. The multiplicity of consistent geometries is taken to provide evidence for this: if those geometries are admitted, then, holding to the standard view of mathematics as interpreted theory, mathematicians will be in the position of asserting manifestly incompatible theories, amongst which mathematics, as such, is powerless to distinguish. But if we accept the Weierstrass/Dedekind/Cantor theory of continuity, then we will have to *accept* this multiplicity of geometries, *at the cost of abandoning the conviction that what is provable in mathematics is TRUE.*

This situation, I surmise, must have seemed to Kronecker the inevitable cost of violations of (**K2**). If we admit the non-constructive theory of limits and continuity, he thought, then the link between mathematical demonstration and truth will be lost. Kronecker's dictum about the divine origin of the natural numbers, and the somewhat more rational remark quoted above about mind as the origin of number, then takes on the characteristics of a *diagnosis* of the unacceptability of this situation. The basic thought, I suggest, is the ancient conviction that the products of the mind alone can be transparent to the mind. The form that this transparency takes in the case of mathematics, for Kronecker, is determined fundamentally by our notion of *effective computability*, the notion that is the basis for our modern understanding of the mathematical import of (**K2**). Genuine mathematics - mathematics that can be determined a priori, without any input from the tainted source of geometrical intuition - must always deal with decidable constructions from the natural numbers, if the link between mathematical demonstration and mathematical truth is not to be broken. The exclusion of geometry, and mechanics, from pure mathematics can then be justified by the observation that nature, which provides the subject matter of those sciences, is not the work of the mind. Consequently, there is no reason to suppose that the structures of the natural world, the structures dealt with in geometry and mechanics, should all be of an effectively computable kind, for there is no reason to believe that we can come to know even the fundamental structure of the natural world a priori. And indeed, those fundamental structures are not knowable in this way: this, for a Kronecker, is precisely what is shown by the arithmetic intractability of the notion of limit.

---

[31] See Kronecker [1887], p253.

Now of course, Hilbert denies (**K2**) - indeed, Hilbert is generally associated with a very radical kind of rejection of (**K2**) and (**K1**) as well, according to which mathematical theories as such are to be thought of as purely syntactic objects. This position is naive *formalism*, so often said to be the official position of the typical contemporary mathematician: mathematics is the theory of syntactic operations on purely formal objects, and therefore mathematical theories as such are neither true nor false. But as we shall soon see, this was not Hilbert's position. Hilbert was never a formalist in this sense, and his arguments against (**K2**) have a far subtler and more interesting character than those available to this naive kind of formalism.

Before we turn to those arguments, though, we shall have to consider the second front Hilbert was defending, this time against the attacks of Frege.

Kronecker, like Hilbert himself, was a creative mathematician, and not a philosopher. Frege, on the other hand, whilst not a creative mathematician of any consequence outside the field he virtually created - mathematical logic - was a very great philosopher indeed. In the famous dispute with Hilbert over the nature of geometry, Frege had little difficulty in exposing the confusions and obscurities in his eminent rival's position. Frege read Hilbert's *Die Grundlagen der Geometrie* on its original publication in 1899, and shortly afterwards studied the lecture notes on which the printed text was based. A brief correspondence with Hilbert ensued, to which Hilbert contributed little. Frege's contribution to the controversy continued for some time after Hilbert's withdrawal, culminating in the two long articles on the foundations of geometry that are amongst Frege's last published writings.

The Frege-Hilbert correspondence has been much commented upon in the literature, and have little to add to the discussion. Most commentators, however, have used the correspondence to shed light on Frege's philosophy of logic, many details of which are indeed trenchantly displayed in his lecturing of Hilbert. Frege's views on geometry, on the other hand, have generally been neglected. But this controversy is not about the philosophy of logic: it is about the philosophy of mathematics, and of geometry in particular. And with respect to this aspect of the controversy, there are some things that need to be said.

Frege's reaction to the multiplicity of consistent geometries was no less drastic than that of Gauss and Kronecker, since his adherence to (**K1**) was no less unqualified than that of Kronecker. In his published writings on this topic, Frege is circumspect; but his unpublished writings leave no doubt about what his views were. For Frege, Euclidean geometry, and Euclidean geometry alone, is genuine science and indisputably part of pure mathematics. The alternatives to Euclid, for Frege, are quite simply tissues of falsehoods. Indeed, he went so far as to compare non-Euclidean geometries with alchemy and astrology - mere pseudo-sciences, with no genuine cognitive value whatsoever:

> If Euclidean geometry is true, then non-Euclidean geometry is false, and if non-Euclidean geometry is true, then Euclidean geometry is false.
>
> If given a point not lying on a line one and only one line can be drawn through that point parallel to that line then, given any line l and point P not lying on l, a line can be drawn through P parallel to l and any line that passes through P and is parallel to l will coincide with it.
>
> Whoever acknowledges Euclidean geometry to be true must reject non-Euclidean geometry as false, and whoever acknowledges non-Euclidean geometry to be true must reject Euclidean geometry.
>
> People at one time believed they practised a science, which went by the name of alchemy; but when it was discovered that this supposed science was riddled with error, it was banished from among the sciences. . . . The question at the present time is whether Euclidean or non-Euclidean geometry should be struck off the role of the sciences and made to line up as a museum piece alongside alchemy and astrology. If one is content to have only phantoms hovering around one, there is no need to take the matter so seriously; but in science we are subject to the necessity of seeking after truth. There it is a case of in or out! Well, is it Euclidean or non-Euclidean geometry that should get the sack? That is the question. [32]

Characteristically, Frege is also completely explicit and forthright about what it is that determines the correct interpretation of the geometric primitives. It is not physical space: the decision between Euclidean and non-Euclidean geometries is not to be made by physics. Truth here is determined by what he calls 'the geometrical source of knowledge'. This is Frege's version of mathematical intuition.

In virtue of his acceptance of (**K1**), the epistemological and cognitive status of the axioms of an axiomatized geometrical theory seemed problematic to Frege. On his view, axioms were Thoughts - not expressions of Thoughts, notice, but the very Thoughts themselves. To distinguish Fregean axioms from axioms in the now familiar sense, let us call the

---

[32] Frege, 'On Euclidean Geometry', in Frege [1979] p169 The editors of Frege's unpublished writings are unable to date this fragment more accurately than 1899-1906. However, there seems to be little doubt that Frege held to this view throughout his life - see for example the extensive fragment 'Logic in Mathematics', dated as Spring 1914 in Frege [op cit], especially pp247 ff; and the very late 'Sources of Knowledge in Mathematics and the Mathematical Natural Sciences' of 1924-25 ([op cit] pp 267-274).

former 'Axioms' (where the distinction does not matter, I use 'axioms'). Since Axioms are asserted directly in proofs, rather than on the basis of some inferential relation amongst previously asserted Thoughts, our grasp of the truth of an Axiom must be direct, unmediated by inference. Our acknowledgment of the truth of an Axiom is grounded directly in the content of that particular Axiom, rather than via an acknowledgment of an inferential relation amongst Thoughts.[33] This means that an Axiom, for Frege, must not only be true, it must also have the property of being, so to speak, transparently true, to anyone who is capable of grasping it at all.

The qualification here is essential, however. It is not Frege's view that there will necessarily be unanimity on the truth of an Axiom. His view, rather, is that an absence of unanimity in such a case will always be attributable to imperfect understanding. This is particularly important with respect to the primitive expressions of an axiomatized theory - the geometrical primitives, in the case of geometry. Just as there must be theorems asserted directly on the basis of their evident truth - Axioms, that is - there must also be *concepts* which must be grasped *prior* to the deductive elaboration of the scientific theory in which they are embedded. For the concepts associated with the primitives of the theory - the concept of line, of point, of plane etc. - must all figure in the Axioms of the theory, and if those axioms are to be grasped as immediate truths, those concepts must already be understood.[34] There is then a potentially very difficult and elusive task, on Frege's view, of ensuring that these primitive concepts are made fully perspicuous. This task is accomplished by what Frege calls *elucidation* (erläuterung).

The only developed example of Fregean elucidation to be found in Frege's own writings is his brilliant analysis of the fundamental concepts of arithmetic, given in parts of the

---

[33] Is this claim not incompatible with the Context Principle, though? I do not think so. Frege can acknowledge a degree of holism in the grasp of the primitive concepts of a deductive theory - he can acknowledge, for example, that there is no grasping the concept of point independently of a grasp of the concepts of line, plane, incidence, congruence etc. In this sense, one cannot enquire into the sense of the concept point independently of its occurrence in sentences - in particular, in the axioms of Euclidean geometry. But whatever the Context Principle means to Frege - and this is a vexed question in Frege analysis - it does not have the consequence of enabling him to allow that there was some non-Euclidean notion of straight line, in virtue of which one could truly think the thought that I would express in English by saying that two straight lines may enclose a space. For Frege, at least one such line must be curved.

[34] Notice that all of this is compatible with there being alternative, equally satisfactory axiomatizations of a mathematical theory, since there may be a set T of theorems possessing the epistemic and cognitive characteristics of axioms, all of which are 'trivially' derivable from each of several different choices of subsets of T. In the same way, there may be several different, equally satisfactory choices of primitive concepts for an axiomatized theory, any of which will enable all the others to be formally defined in the theory. Where this is so, there is a 'local holism', as it is sometimes called, amongst the primitive concepts.

Foundations of Arithmetic and subsequent works. Elucidation here appears in the guise of a *reductive philosophical analysis*, which aims to show that the primitives of arithmetic, properly understood, are in fact logical in nature (logical by Frege's lights, of course).[35] Elucidation need not always take this very elaborate form, however - it does so in the case of arithmetic precisely because Frege thinks that the real nature of the primitive concepts of arithmetic is not at all obvious. Frege would have thought of Euclid's notorious 'definitions' of 'point', 'line' etc. as elucidations. He would have taken Cantor's famous remark 'By a *Menge*, we are to understand any collection into a whole M of definite and separate objects m of our intuition or our thought' as an elucidation of the concept of set, and he would have thought of the more recent explanations of the concept of set via a description of the generation of sets by an iterative process in the same light.[36] Evidently, this notion of elucidation is not a very precise one, since what will count as an elucidation will in general depend on a host of unpredictable and pragmatic considerations, such as the audience for whom the elucidation is being offered. Indeed, this imprecision is for Frege an essential characteristic of elucidation, and the primary reason why he insists time and again that elucidation must not be confused with definition. Definition is a scientifically precise activity, which takes place within an axiomatized theory: elucidation is pre-scientific, and must never be thought of as providing genuinely *scientific* grounds for the assertion of any Thought.

With these qualifications duly noted, however, it remains Frege's view that any doubts about an Axiom - as opposed to doubts as to whether some Thought is in fact suited to the role of an Axiom in a deductive theory - are necessarily attributable to misunderstanding. With characteristic forthrightness and integrity, Frege directly acknowledges the implications of this view for the problem for (**K1**) posed by multiplicity of geometries:

> Can we not put to ourselves the question: How would it be if the axiom of parallels did'nt hold? Now there are two possibilities here: either no use at all is made of the axiom of parallels, but we are simply asking how far we can get with the other axioms, or we are straightforwardly supposing something which contradicts the axiom of parallels. It can only be a question of the latter case here. But is must constantly be borne in mind that what is false cannot be an axiom, at least if the word 'axiom' is being used in the traditional sense. What are we to say then? Can the axiom of parallels be acknowledged as an axiom in this sense? When a straight line intersects one of two parallel lines, does it

---

[35] With respect to the primitives of geometry, we do not know in any detail what a Fregean elucidation would have looked like. Perhaps he thought that the groundwork at least of this task had been accomplished by Kant, for he certainly believed that Kant had been correct about the cognitive and epistemic status of geometry. Certainly, his account would not have been reductive.

[36] Cantor [1955], p85. For the iterative conception, see e.g. Boolos [1971], or Scott [1967].

> always intersect the other? *This question, strictly speaking, is one that each person can only answer for himself. I can only say: so long as I understand the words 'straight line', 'parallel' and 'intersect' as I do, I cannot but accept the parallels axiom. If someone else does not accept it, I can only assume that he understands these words differently.* Their sense is indissolubly bound up with the axiom of parallels.[37]

Now of course, on any view, if the formal sentence that expresses the parallels axiom is evaluated as true in one model, and false in another, then the constituent expressions of that formal sentence, i.e. the geometric primitives 'line', 'plane', 'parallel' etc., must be in *some* sense interpreted differently in the two models. This has nothing to do with what Frege is saying, however. Frege is not interested in anything that follows from the possibilities of interpreting formal theories (in our sense) in alternate ways.

Now, there is a perfectly intelligible sense in which two disputants, one of whom thinks that a hyperbolic geometry correctly characterizes physical space, whilst the other thinks that Euclidean geometry correctly characterizes physical space, understand the geometric primitives differently. Each of these disputants can characterize his understanding of the geometric primitives to their mutual satisfaction, and the dispute between them is a dispute over which of these mutually intelligible geometric systems applies in physical space. Even if we drop the reference to physical space, and adopt (say) a set-theoretic interpretation for geometric systems, there is still a perfectly intelligible sense in which the geometric primitives are 'understood' differently as they are interpreted over set theoretic models of the alternative consistent geometries. But Frege takes himself to be able to discriminate a meaning for the geometric primitives which guarantees, not the truth *in a model* of the formalized parallels axiom, but the *truth*, simpliciter, of the Thought that is the parallels Axiom in his sense.

But what can be said to those who doubt that this Thought is true? Well, what does it *mean* to doubt that this Thought is true? For example, ought our judgment here turn on whether physical space is Euclidean? Clearly not: Frege does not think that the truth of his parallels Axiom is an empirical matter. No: the claim is that there are senses, accessible independently of this or that scientific geometric theory, and independently of how things actually are with space, which are such that, if one has grasped them correctly, one cannot but grasp the parallels Axiom as a truth. This is the claim.

---

[37] Frege, 'Logic in Mathematics', in Frege [1979] p247, emphasis mine.

Since the question of the truth of the parallels postulate, according to Frege, is something that one can only judge for oneself, the response to this has to be given in the first person. For my part, I have to say that these senses have eluded me. With hand on heart, I have to say that the Euclidean parallels axiom does not strike me as obviously true - for reasons that were already troubling to Proclus.[38] Perhaps some educational programme would help: further study of the Critique of Pure Reason, maybe, or Frege's own never discharged obligation to elucidate the sense of the geometric primitives. But I suspect that the educational project would fail, if only because so many men have undertaken just this project in the past, and failed. Faced with a genuine case of such a failure, Frege's philosophy of mathematics is dumb. There is simply nothing to be said, other than that I, or he, must have failed to understand. It seems to me, as it surely did to Hilbert, that this is an intolerably solipsistic basis on which to place the foundations of the mathematical sciences.

Of course, I can indeed associate meanings with the primitives of Euclidean geometry in such a way as to ensure the truth of the Thought that a line which intersects one of two parallel lines must intersect the other (say). I can do so precisely by allowing the Euclidean primitives to have whatever meanings are necessary to ensure the truth of the axioms of Euclidean geometry, from which this proposition is derivable. This position, though, is not available to Frege, for now the sense of the geometric primitives is given to me *through an axiomatized geometric theory*, and not independently of it. In exactly the same way, I can attach meanings to the geometric primitives in such way as to ensure the truth of the Thought that parallel lines intersect at infinity, now allowing the geometric primitives to have whatever meaning is necessary to ensure the truth of the axioms of projective geometry, say. In this way, I can come to understand perfectly well geometries in which the parallels axiom takes opposed truth values, and I am willing to say that the geometric primitives have different meanings in these alternative systems. But if someone then asks me, which of these interpreted geometries is *TRUE*, I have to ask for some further clarification before I can answer. If the question concerns physical space, then I will have to defer to a physicist for judgment. If the question turns upon the mathematical acceptability, in any sense I can understand, of the two systems, I shall have to reply that, so far as mathematics is concerned, there is no reason to deny that they are both true. If the question is intended to resolve Frege's question, though - the question about the TRUTH of the parallels Axiom - I have to reply that *this* question simply has no answer.

---

[38] See Gray [1989] pp34-36.

For consider the alternative, Fregean response, in the light of the actual history of the Euclidean parallels postulate. The single most striking feature of that two thousand year long dialectic, in which proof after proof of the parallels postulate was entertained and rejected, is the complete inability of the mathematical community to agree on what counts as a self-evident truth, and what counts as an assumption that requires demonstration. (This has nothing to do with the famous 'hidden assumptions', incidentally. Most of them were not hidden: they were regarded as too obvious to merit explicit acknowledgement.) Wallis's famous proof makes the (explicit) assumption that, for any given triangle, congruent triangles exist of every size. Is this self evident? It can in fact be shown to be deductively equivalent to the parallels postulate, which of course robs Wallis's proof of any suasive power. But the point here is that this congruence 'Axiom' is something which struck Wallis, and many others, as obvious, whilst the Euclidean parallels Axiom struck them as requiring proof; whilst for others, of course, the situation was exactly reversed. This deadlocked pattern of futile attempts at persuasion repeats itself over and over again with an enormous number of 'Axioms' which turn out to be equivalent to the parallels postulate; and throughout it all there is a complete, settled inability of competent mathematicians to come to any agreement on which, if any, of these 'Axioms' neither had, nor needed, proof. And on the other hand, the inquiries which in fact generated non-Euclidean geometries - that is to say, the attempts by Saccheri et al to prove the parallels postulate by a reductio argument from its negation - were similarly deadlocked by an inability to get any consensus on what constituted a genuine reductio ad absurdam. One man's reductio, after all, is apt to be another's non-Euclidean geometry. It stretches credulity too far to think that there is some understanding of Euclidean geometry which Frege had, but which these great mathematicians lacked, which is such that it would have sufficed to set their enquiring minds at rest. Nor need we turn to ancient history for examples of this kind of futility. Amongst the many known equivalents of the Axiom of Choice, a mathematician chosen at random from the mathematical community of the 1920's, say, would be apt to find some obvious, some less obvious, and some hard to swallow. But which versions fell into which categories is something that might vary widely, depending upon your choice of mathematician. Amongst the many who inveighed most loudly against the axiom on its first introduction, there were very few who did not rely, constantly but unconsciously, upon some equivalent of the axiom of choice in their own work.[39]

---

[39] The paradigm case is Poincare - see Moore [1982].

Of course, Frege was not unaware of the implications of his position. On the contrary, he understood them very well. Sometimes puzzlement is expressed as to how Frege could still believe these things when it was already widely held amongst physicists that physical space was non-Euclidean, and I have heard speculation to the effect that Frege may not have kept sufficiently in touch with physics to be aware of this. That is surely false. Frege was a student of physics as well as mathematics, and remained professionally interested in physical theory as a practising mathematician. Frege's commitments with respect to geometry are not the consequence of ignorance of physics, or of any lack of awareness of the history of geometric theorizing. They are consequential upon his understanding of the nature of logic, of inference, and especially upon the powerful, sophisticated and highly plausible semantic theory that he had created. It is this body of doctrine, in conjunction with his adherence to (K1), that forces these unpalatable conclusions upon him. To give them up, he would have had to abandon some central tenets of his philosophy.

Section Four: Hilbert and the Philosophy of Mathematics. I think that it will help us to understand Hilbert's work on the foundations of mathematics if we see it as a self-conscious attempt to find a middle way between the revisionist extremes represented by Kronecker and Frege. For this perspective reveals the deep continuities underlying the early work in geometry and the later work on the consistency of classical mathematics, and a sensitivity to this underlying continuity is the key to a correct understanding of Hilbert's Programme.

Consider, for example, the question of why Hilbert should ever have written *Die Grundlagen der Geometrie* at all. Occasionally, one hears that the cardinal virtue of Hilbert's axiomatization is that it reveals the 'hidden assumptions' on which the validity of a great many Euclidean proofs depends, and thus creates a system of Euclidean geometry which really does have the deductive rigor mistakenly attributed to Euclid's own work. But why should Hilbert have wanted to do that? The 'hidden assumptions' were all known before Hilbert's axiomatization appeared, and Hilbert had no interest in rigor unless it served some clear mathematical purpose. Euclidean geometry was hardly a flourishing research programme in the late nineteenth century, and the primary interest of Hilbert's axiomatization does not lie in any wealth of new and unexpected results. What, then, was the point? Why should this ambitious young research mathematician, with a brilliant reputation but an as yet unestablished career, devote so much time to working over an apparently exhausted vein?

The answer is that *Die Grundlagen der Geometrie* is primarily a work on the foundations of mathematics, not geometry. More precisely, it is aimed at exemplifying, clarifying, and popularizing a certain foundational programme - the embryonic form of Hilbert's Programme. Now, this description is apt to seem surprising, since Hilbert's Programme is dominated by the search for finitary consistency proofs, whilst in the *Grundlagen*, concern for the consistency of geometric systems is not nearly so prominent, and is satisfied by a model-theoretic argument establishing the consistency of his axiomatization of Euclid's geometry relative to a theory of the real numbers. This seems very remote from Hilbert's Programme. If we look more closely, though, the underlying continuity of concerns becomes apparent, and with it the origins of Hilbert's later obsession with finitary consistency proofs.

Seen in the light of the difficulties in Frege's position, the aspect of Hilbert's *Grundlagen* which will strike us most strongly is the determined attempt to shift the discussion of geometric systems away from extra-systematic questions concerning the epistemic and cognitive status of axioms and primitive concepts, towards 'internal', systematic questions concerning derivability, independence, completeness, and consistency. Yet Hilbert calls this a work on the *foundations* of geometry. Frege's work on the foundations of arithmetic, remember, took the form of a reductive elucidation of the arithmetic primitives, and he expected to find Hilbert at least addressing similar questions with respect to the geometric primitives. But Hilbert did no such thing. Instead, he stated a series of axioms, and then claimed that those axioms themselves constituted a *definition* of the geometric primitives they contained. Frege was incensed. He found himself faced with a work on the foundations of geometry which, as he said, left him unable to determine whether his pocket watch was a point. Since he took it to be mortally certain that his pocket watch was *not* a point, this struck Frege as conclusive evidence that Hilbert's work was gravely defective.

Yet ten years before writing the Foundations, Hilbert remarked in conversations on geometry: "One must be able to say at all times - instead of points, straight lines, and planes - tables, chairs, and beer mugs" - or indeed, pocket watches.[40] As he put the point in his only substantial contribution to the exchange of letters:

> . . . you say that my concepts, e.g., 'point', 'between', are not unequivocally fixed. . . . *But it is surely obvious that every theory is only a scaffolding*

[40] See Reid [1986], pp57-64.

> *(schema) of concepts together with their necessary connections, and that the basic elements can be thought of in any way one likes.* E.g., instead of points, think of a system of love, law, chimney-sweep . . . which satisfies all axioms; then Pythagoras' theorem also applies to these things. Any theory can always be applied to infinitely many systems of basic elements. For one only needs to apply a reversible one-one transformation and than lay it down that the axioms shall be correspondingly the same for the transformed things (as illustrated in the principle of duality and by my independence proofs). All statements of electrostatics hold of course also for any other system of things which is substituted for quantity of electricity . . ., provided the requisite axioms are satisfied. Thus the circumstance I mentioned is never a defect (but rather a tremendous advantage) of a theory.[41]

The stage was not set for a meeting of minds.

And yet, in the end, there was indeed something not too far from a meeting of minds, although it remained unacknowledged by both Frege and Hilbert.[42] Reading Frege's contribution to the exchange nowadays, one feels a mixture of admiration and exasperation as Frege painstakingly translates what he (quite rightly) saw as Hilbert's deeply confused presentation into his own terminology, eventually to come up with the conclusion that, if Hilbert had succeeded in defining anything at all, then he had in fact defined (what we would now call) a Euclidean structure. And of course, behind all the confused talk of axioms as 'implicit definitions', that was *exactly* what Hilbert was trying to do.[43]

But there are revealing differences between Frege's reconstruction of Hilbert's geometry, and Hilbert's original. Frege noted that Hilbert's axioms contained, in addition to first order predicates such as 'x is a point', 'y is a line' etc., certain second order predicates - or, in Fregean terms, quantifiers. If we view the first order predicates as variables, we can then put them in the argument places of the second order predicates, and in this way construe the conjunction of Hilbert's axioms as a single, second-order relational predicate - and this is at least akin to what we would now call a Euclidean structure.[44]

But for Frege, of course, the second order predicates must themselves be taken to be implicitly bound by third order universal quantifiers, whereas Hilbert would have wanted them to be thought of as schemata, in accordance with the approach of modern model

---

41 See Frege [1980], pp42-43

42 So far as I know, Hilbert never comments on his exchange with Frege on the matter of axioms and definitions in his later writings. However, Bernays takes up the topic in his review of a published version of the correspondence, and acknowledges that Frege was in the right. See Bernays [1942].

43 Frege also came to see Hilbert's motivations correctly - see the beginning of his second letter to Hilbert, where he notes Hilbert's attempts to free geometry from intuition (Frege [1980], p43)

44 Cf. Resnik [1974].

theory.[45] Nevertheless, Frege felt that he had uncovered what was really going on in Hilbert's confused presentation, not least because the generality of application that Hilbert evidently wanted - not just points, but pocket watches - was now accommodated in terms Frege could understand. As Frege has reconstructed it, generality has been secured, not by varying interpretations of schemata constructed from constant expressions of indeterminate sense, but by instantiation of universally quantified variables over a fixed, perfectly determinate domain. There is no theory here which can be interpreted in alternate, equally acceptable ways - no notion of truth in a model. Rather, there is a theory with a completely fixed interpretation at a higher level, and its axioms and theorems must be TRUE.

Hilbert has been rescued, then, at the price of inflicting upon him the Fregean conception of generality, shown to be of doubtful coherence by the paradox Russell discovered in the system of the *Grundgesetze*. There is a double misrepresentation of Hilbert's intentions here, for Hilbert has been burdened with a kind of generality he did not want, and simultaneously denied the kind of generality he did want. Frege's logical theory is a theory of types, with all of its variable expressions and quantifiers appropriately stratified. Hilbert's schemata, however, are not so stratified, and they are not intended to be interpretable by expressions of some independently determined, fixed type. Nothing expressible in Begriffschrift can have *this* kind of generality. Consequently, the reconstructed Hilbert is faced with the task of determining the fixed domain of interpretation for the various types of variables involved in Frege's formalization of his theory, which is exactly the kind of extra-mathematical task that Hilbert approach aimed to avoid.

Consequently, even from the vantage point of his charitable reconstruction of Hilbert's project, Frege can make very little sense of the kind of systematic questions to which Hilbert devotes such energy in the *Grundlagen*. Paradigmatic here is Frege's perplexity over Hilbert's interest in independence and consistency proofs. Frege could make some sense of the idea of an independence proof for an axiom, since in his reconstruction, one gets such a proof by specifying a sequence of first-level concepts such that, with those concepts instantiated at the argument places of the second order quantifiers, the distinguished axiom becomes false, and the remainder of the axioms true. And this, again

---

[45] This is potentially misleading, since Hilbert, at this time, was not aware of the need to distinguish the semantic notion of logical consequence from the notion of derivability in a formal system. Thus, whilst he does indeed apply something very like a contemporary model theoretic approach to questions of consistency, independence etc. in the Grundlagen, it is also clear that he thinks of logical consequence purely in terms of derivability from the axioms of geometry.

in an unfamiliar terminology, resembles the procedure Hilbert actually used in proofs of independence in the *Grundlagen*. Of course, there is nothing here that is incompatible with the evident truth of the Euclidean parallels Axiom, since nothing in this procedure, as Frege has reconstructed it, suggests fixing a sense for a hitherto senseless, or in some way semantically indeterminate expression.

For this same reason, Frege is unable to make anything of the request for a proof of consistency. To us, proofs of consistency and proofs of independence are intimately related, since a model theoretic consistency proof in effect shows that some sentence - $\ulcorner 0=1 \urcorner$ say - is independent of the axioms. On Frege's reconstruction, though, this is always trivial. You just leave everything as it is - you apply the identity transformation to the Axioms and to $\ulcorner 0 = 1 \urcorner$. There simply are no Axioms which are false, no absurdities which are true, and this is transparent once the translation of a theory into Begriffschrift is completed. Hilbert, of course, would not have been sanguine about the prospects for a translation of Euclidean geometry, or any other mathematical theory for that matter, into Begriffschrift. For that required the prior elucidatory task of fixing the senses of the geometric primitives, and the conclusion Hilbert drew from the history mentioned above was that any philosophy of mathematics that demanded successful completion of *that* task left mathematics completely at the mercy of what he called 'the inadequate means of metaphysical speculation'.

For the lesson that Hilbert, along with the great majority of mathematicians, drew from the history culminating in the discovery of a myriad of consistent alternatives to Euclid, was precisely that Frege's prior task of determining the sense of the geometric primitives was impossible for mathematics to discharge. So far as Hilbert was concerned, all of the mathematically manageable meaning of the geometric primitives had been shown to be contained within what we would now call the *logical structure* of the theory, for it was only with respect to what was determined by the logical structure that the mathematical community had been able to achieve any kind of consensus.

Unfortunately, Hilbert initially resisted this Fregean demand for some prior determination of the sense of the geometric primitives by advancing a very strong rejection of the principle (**K1**), implicit in the following extract from his one lengthy letter to Frege:

> I was very much interested in your sentence: 'From the truth of the axioms it follows that they do not contradict one another', because for as long as I have been thinking, writing, lecturing about these things, I have been saying the

> exact reverse: *If the arbitrarily given axioms do not contradict one another, then they are true, and the things defined by the axioms exist.* This for me is the criterion of truth and existence. . . . It is precisely the procedure of laying down an axiom, appealing to its truth [this is its intuitive truth, in the Fregean sense], and then inferring from this that it is compatible with the defined concepts that is the eternal source of errors and misunderstanding.[46]

On the face of it, this is an extreme version of mathematical *realism*, a realism which is intended to be completely free from reliance on mathematical intuition. We might encapsulate it in the following principle:

(H1) If a set S of mathematical sentences is consistent, then the objects mentioned in or quantified over in the sentences of S exist, and the sentences of S are true.

Now, Frege's initial response to **(H1)** seems to me quite inept, for he asks if the consistency of the set of sentences S = {'A is an intelligent being', 'A is omnipresent', 'A is omnipotent'} licenses an inference to the existence of an intelligent, omnipresent, omnipotent being. But **(H1)** does not threaten to license that inference, for S is not a set of mathematical sentences.[47] And Hilbert is quite clear that **(H1)** has no plausibility whatsoever outside of mathematics. In the case of physics (or natural science in general) where Hilbert was just as inclined to press the demand for axiomatization, he never makes the (ludicrous) suggestion that any consistent set of arbitrarily selected axioms has a model.

Still, there is a better worry lurking behind this question, for it now appears that Hilbert at least owes us some independently drawn distinction between mathematical and non-mathematical sentences, if **(H1)** is to hold only for the former. Given the inability of Hilbert's 'implicit definitions' of the geometric primitives to settle the issue of whether Frege's pocket watch was a point, it might be thought that no such account could be forthcoming.[48] In effect, this is the second complaint Frege raises against **(H1)**, when he asks if there is any way of demonstrating inconsistency 'besides pointing out an object that has all the properties'.[49] This cuts much deeper. For if the only way to establish the

---

[46] Frege [1980] p42, emphasis mine.

[47] Nor, for that matter, is S obviously consistent.

[48] Not that Frege is in any better shape with respect to an answer to this question. For the issue that has now been raised is the notorious Caesar problem of *Die Grundlagen der Arithmetik*. And the reader will recall that it is in response to the Caesar problem that Frege makes his fatal appeal to extensions, thus inducing the inconsistency in his system discovered by Russell. Hilbert, incidentally, thought that this kind of failure was symptomatic of the futility of any attempt to discriminate the subjects matter of mathematics externally - by way of a metaphysical, rather than a mathematical, argument. He thought it no coincidence that Dedekind's famous 'proof' that a simply infinite system exists founders in a closely analogous way.

[49] See Frege [1980], pp47-48.

consistency of a set of sentences is by 'pointing out' some objects satisfying the predicates of those sentences, the claim that consistency provides a 'criterion' of truth and existence in mathematics evidently gets things exactly the wrong way round.

Hilbert is now in considerable difficulties. His own consistency proofs, in the *Grundlagen*, have been model-theoretic in character. He has shown the consistency of his axioms for Euclidean Geometry precisely by 'pointing out' some objects satisfying the predicates of those axioms - by interpreting geometry in the real numbers. But where is the 'criterion of truth and existence' for the axioms 'defining' the real number system to come from? A model in set theory? It is now very tempting to conclude that, sooner or later, a proof of consistency is going to have to be given directly, without any appeal to models, if Frege's regress is to be halted. Here we have one principal impetus for Hilbert's later search for purely syntactic proofs of consistency - and it is to be noticed that it originates in an attempt to provide an account of truth and existence in mathematics.[50]

Let us back away from the specific problems posed by (**H1**) for a moment, to get clearer about the pressures that are shaping Hilbert's responses here. One of the principle convictions that is animating Hilbert is the belief that there are no mathematically manageable grounds for discriminating amongst the various consistent geometric systems with respect to intelligibility, meaningfulness, truthfulness, or whatever. Any attempt to fix the One True Geometry, aside from abandoning geometry to physics, must either take the form of introducing yet another axiom system, in which case the problem posed by consistent alternatives remains untouched, or else take on the hopeless Fregean task of fixing the sense of the geometric primitives independently of mathematical investigation. Now, Hilbert's real target here is any mathematically revisionary position - such as Frege's, or Kronecker's - which seeks to distinguish between various consistent mathematical theories, admitting *some* to the canon of pure mathematics and excluding

---

[50] This is why it is worth pointing out that (**H1**) is not an expression of formalism: it is an expression of a strong form of realism. Penelope Maddy, in her recent book on realism in mathematics, describes the early Hilbert as a deductivist (what she calls an 'if-thenist'). Now, deductivism is the view that mathematicians simply explore the deductive consequences of uninterpreted axioms - it is indeed a kind of formalism. I cannot find this view anywhere in Hilbert, early or late. On the contrary, the Hilbert of the *Grundlagen* and the correspondence with Frege is plainly attempting to find a way of defending the objective truth of classical mathematics, including hyperbolic geometries. To be sure, he does so by way of a quite inept attempt to combat Frege's highly sophisticated understanding of the notions of truth and existence in mathematics; but nevertheless, his attempt does *not* take the form of a denial that the concepts of truth and existence have any application in mathematics. (Ironically, Maddy's own version of mathematical realism comes perilously close to adopting (**H1**) as an account of mathematical truth and existence - as we shall see in the following section.)

others, on grounds which are themselves external to mathematical theory. But in arguing against revisionism of this kind, revisionism which is inspired by argumentation which is philosophical, or semantical, or theological, rather than properly mathematical in character, Hilbert is always inclined to sound as if he is advancing a position according to which mathematical theories are *meaningless*.[51] This is why Hilbert is so often thought of as a formalist. However, careful reading will suggest a more sympathetic interpretation, on which he is in effect arguing that *no* mathematical theories are meaningful *in the way that Frege* (or Kronecker, or, later, Brouwer) *thinks*. His claim against Frege is: *not even Euclidean geometry* meets Frege's standards for mathematical meaningfulness (witness the history of the parallels postulate).

But this need not be, and should not be, seen as an attack on the meaningfulness of mathematics: it is an attack on a particular, *philosophically motivated and mathematically revisionary* conception of what is required for mathematics to be meaningful. Hilbert wants to resist any revisionary conception of what is required for the meaningfulness of mathematics which rests upon considerations which cannot themselves be determined mathematically, which puts mathematical truth at the most fundamental level beyond the reach of what is mathematically demonstrable.

Now, the kind of mathematically revisionary philosophical conception of what is required for the meaningfulness of mathematics adopted by Frege and Brouwer is to be sharply contrasted in this respect with the position taken more recently by Hartry Field. A consequence of Field's philosophical position is that mathematical assertions are never (non-vacuously) true (or false): mathematics is semantically defective, but this defect is a feature of *all* of mathematics, and has no implications whatsoever for the professional practice of mathematicians. Field is not a revisionist about mathematics: he mounts no philosophical attack on any established mathematical theory, and he favours no mathematical theory over any other with respect to meaningfulness.

This is a very different kind of philosophy of mathematics than anything Hilbert ever contended against. To take issue with Field, one really is required to contest several related

---

[51] The fact is, though, that he typically speaks of his 'ideal elements', not as meaningless, but rather as having no meaning apart from their role in an axiomatized mathematical theory. This provides some textual justification for the attempt, which I am now in effect making, to show that a kind of use-based theory of meaning comports rather better with Hilbert's overall position that any kind of formalism does. Genzen, for example, interprets Hilbert in this way - see e.g. Genzen [1938], English translation in Genzen [1969], especially pp247-251.

doctrines in the metaphysics and the philosophy of language - doctrines concerning the causal theory of reference, the intelligibility of discourse about abstract objects etc. This is a straightforwardly and explicitly philosophical controversy, and so far as I can see, Hilbert need have nothing to do with it. Given that nothing in mathematics is at stake, given that no part of mathematics is under attack, Hilbert, it seems to me, can simply look the other way.

However, Hilbert, in his dispute with Kronecker and Frege (and later with Brouwer), evidently felt it necessary to engage in just the kind of wider, more properly philosophical dispute about mathematics that Field's instrumentalism invites, but which greatly exceeds anything strictly necessary for the task of defeating Fregean or intuitionistic revisionism.[52] Rather than sticking to exposing the inadequacies of the revisionary positions he is resisting, Hilbert felt obliged to offer some competing, positive philosophical doctrine of his own - and thus we return to the principle (**H1**), together with a number of related theses concerning the existence of the objects of mathematics, and what is required for their existence. These theses, in my view, do not show his thought at its strongest.

Indeed, I do not think that Hilbert has a plausible, detailed philosophical proposal concerning the general issue of the existence of mathematical objects.[53] But that is *not* to say that all of Hilbert's claims about the philosophical importance of the axiomatic method, and in particular about its implications for ontology, are unimportant. In fact, I think that much of what he says on these subjects is interesting and worth taking very seriously - and not just with a view to understanding Hilbert's foundational programme better. And in particular, I think that a sympathetic appraisal of what Hilbert actually says, *together with attention to what he actually tries to do in foundations of mathematics*, will show that his instincts, if not his arguments, were very much on the mark. If we begin now to pay closer attention to what Hilbert actually does in his foundational work, we shall begin to get a clearer view of his attitude to Kronecker, and with it the outlines of a more plausible alternative to (**K1**) than (**H1**).

---

[52] And this, of course, is admirable. Would that it were always true that the best scientists showed such interest in philosophical questions concerning their discipline. However, it seems to me worth pointing out that the philosophical ambitiousness of some of Hilbert's claims greatly exceeds his immediate needs. The point is that most of what is really important to him is available for a much more modest philosophical investment.

[53] Hilbert and Bernays discuss the axiomatic method in the late *Grundlagen der Mathematik*, their approach is pretty much the approach of modern postulation theory (see Church [1956] pp317-332). The talk of axioms 'implicitly defining' their domains has been dropped. In Bernays [1941], there is an explicit acknowledgment that Frege's criticisms of Hilbert's writings on this question had been completely justified.

Kronecker's complaint against geometry, you will recall, turned upon his animadversions against the Weierstrass/Dedekind/Cantor theory of continuity, which violated the finitary strictures imposed by (K2). Hilbert's response to this, in the *Grundlagen*, is to provide an axiomatization of basic Euclidean and projective geometry in which the continuity requirements are isolated and identified with the role played in the overall theory by the two axioms of Group V (the Archimedean axiom, and the so-called Axiom of Line Completeness - the remaining Hilbertian axioms deal with essentially non-numerical ordering principles). At first glance, these continuity axioms seem too weak to support the full theory of continuity attacked by Kronecker, but Hilbert goes on to demonstrate that a geometry equivalent to 'ordinary analytic geometry' can in fact be obtained from this basis:

> **The completeness axiom is not a consequence of Archimedes' Axiom.** In fact in order to show with the aid of Axioms I - IV that this geometry is identical to the ordinary analytical 'Cartesian' geometry Archimedes' Axiom by itself is insufficient . . .. However, by invoking the completeness axiom, *although it contains no direct assertion about the concept of convergence*, it is possible to prove the existence of a limit that corresponds to a Dedekind cut as well as the Bolzano-Weierstrass theorem for the existence of condensation points, whereby this geometry appears to be identical to Cartesian geometry.[54]

What Hilbert has done, in effect, is to show that one can meet the continuity requirements of 'ordinary analytic geometry' without going through the full Weierstrass/Dedekind/Cantor theory of continuity - indeed, without speaking directly of the notion of convergence at all. And in addition to this, Hilbert also shows that a large part of ordinary analytic geometry is independent of the continuity axioms altogether.

The point to emphasize here is that Kronecker's complaint has been addressed by giving a mathematical elaboration of the *weakest notion of continuity* needed for analytic geometry, by way of an axiomatization that renders the exact continuity assumptions needed, and the minimal role they are required to play, completely transparent. Now of course, the result is *not* a system of analytic geometry that meets Kronecker's finitary constraints, the constraints imposed by (K2). Nevertheless, the fangs of (K2) have been drawn somewhat, by both (a) showing that the axiomatic elaboration of analytic geometry is consistent, relative to a theory of the real numbers, and (b) showing that the excess over finitary mathematics involved in analytic geometry is less than one might have supposed. Whilst this will not completely satisfy the determined finitist, Hilbert has nevertheless

---

[54] Hilbert [1971], p28, my italics.

provided a valuable response to the legitimate worry that animates the finitary standpoint. For Kronecker, remember, the deep worry behind (**K2**) was a worry about the acceptability of *conceptual innovation* in mathematics. What constrains the introduction of new apparatus in attempts to solve familiar problems? According to (**K2**), the constraint must be, *finitary reducibility to the natural numbers*. Unable to accept the drastic consequences of this, Hilbert's first 'philosophical' alternative to (**K1**) is (**H1**): *consistency* is the sole constraint. However, Hilbert's *actual practice* - early and late - suggests that he had something more interesting in mind: a demonstrably consistent axiomatization of the new apparatus, *but also a careful investigation, using minimal means, of the role that it plays with respect to the mathematics in which the problem originates.*

It is instructive to compare this with one striking feature of the mature Consistency Programme of the 1920's. The crucial issue here concerns the way in which, according to Hilbert, existence theorems are to be proved in axiomatized mathematical theories. Now, in the mature form of Hilbert's logical theory (from about 1925 onwards), the quantifiers are not primitive symbols. Rather, they are introduced as defined symbols by the use of Hilbert's 'logical ε-axiom':

(ε) $A(a) \rightarrow A(\varepsilon(A(x)))$

The ε-symbol used here is interpreted as a kind of choice function (syntactically, it is a term forming operation on predicates). In Hilbert's own revealing explanation, given any property F(x), this function picks a paradigm case of F if anything satisfies F(x), and otherwise chooses at random.[55] The quantifiers may then be defined as follows

(∀) $(\forall x)F(x) \leftrightarrow F(\varepsilon(\neg F(x)))$
(∃) $(\exists x)F(x) \leftrightarrow F(\varepsilon(F(x)))$

And just as Hilbert's axiomatization of geometry sought to clarify the role played by continuity in geometry by isolating the deductive roles played by the two axioms of Group V, so too Hilbert's axiomatization of logic attempts to clarify the role played by the infinite in mathematics in general by isolating the deductive role played by the principle (ε) -

[55] This is revealing, I think, because it suggests once again that Hilbert's conception of generality was naturally expressed in schemata, rather than quantified variables. Thus in explaining how the universal quantifer is to be defined from this choice function, Hilbert offers the following analogy: if even Aristides, the paradigmatic just man, is unjust, then everyone is unjust.

because, as Hilbert insists, *it it via quantification that the infinite enters mathematics*. By isolating the ε-axiom and investigating the role that it has to play in an axiomatized mathematical theory, Hilbert is trying to render transparent the weakest notion of infinity that number theory requires, for the ε-terms are the sole ideal elements in Hilbert's axiomatizations of number theory. Hilbert and his assistants then succeeded in proving a theorem (the second ε-theorem) which does indeed show the possibility of *eliminating* ε-terms from proofs of formulae that do not themselves contain occurrences of any ε-terms. This theorem is the paradigm of Hilbertian elimination of ideal elements. With the second elimination theorem in hand, Hilbert is able to show that, if the system proved something of the form $(\exists x)F(x)$ with matrix containing no ε-terms, then the system must also prove $F(a)$ for some term a recoverable (in principle) by the process which eliminates ε-terms. (It may help to observe here that, in Hilbert's system **Z**, the ε-operator is in fact analogous to the familiar least number operator of recursion theory.) This theorem therefore provides a perspicuous example of establishing the eliminability of ideal elements from proofs.[56]

Now, the process of eliminating ε-terms in some formula(s) φ from some proof P in this fashion induces, in effect, an interpretation *of the ε-terms occurring in P*. But this 'interpretation' need have nothing of semantic significance in common with the interpretations of φ induced by the elimination procedure as applied to any other proof in which φ occurs. Beyond the elimination algorithm itself, there is really nothing more to be said about the interpretation of ε-terms - this is why Hilbert shows no interest in the semantic properties of the ε-symbol.[57] Here, I think, one gets a clear sense of the deep grounds of Hilbert's opposition to both Frege and Kronecker. Against the Fregean demand for a preliminary explication of the sense of the ε-axiom, Hilbert's position is in effect that, beyond the elimination algorithm and the metamathematical demonstration that it works, there is nothing in general to be said that is of any real semantic significance. (If you like slogans, meaning here is just use.) And against the radical reductionism of Kronecker's Programme, implicit in **(K2)**, Hilbert has shown that ideal elements can be incorporated *in a finitarily responsible way* without accepting general reducibility to purely finitary mathematics - since the algorithm only shows how to remove ε-terms from proofs

---

[56] The first ε-elimination theorem establishes the eliminability of the quantifiers in favour of the ε-symbol. As one would expect, the complications arising from the need to extensively relabel bound variables are considerable, and perhaps partially account for the unpopularity of the ε-symbol in the logical literature. For more details, see Leisenring [1969].

[57] I owe this observation to Warren Goldfarb.

of finitary theorems, and in any case does not suggest that there is anything like an interesting, unitary finitary meaning to be associated with the ε-operation.[58, 59]

What emerges from all this, then, is the outline of a proposal for the clarification and justification of the mathematician's use of the actual infinite which seems to me to be of great interest and subtlety. Against the Fregean demand for a preliminary, philosophical explication of the concepts involved in infinitistic mathematics, Hilbert is proposing an internal mathematical investigation of the roles played by this, that or the other infinitistic notion in axiomatized mathematical theories - and in particular, in axiomatic analysis. Against the Fregean complaint that this approach is unprincipled - in response, that is, to the Fregean question, *With what right* to you introduce these infinitistic notions - Hilbert is proposing that the only scientifically manageable way of justifying the use of a concept, of explaining its sense, is by giving a careful account of the role that the concept actually plays within an axiomatized theory which meets overall standards of mathematical acceptability. And against the Kronecker complaint that this approach threatens to trivialize mathematics, by permitting arbitrary innovations without regard for the conceptual scheme within which the particular mathematical problems requiring solution have arisen, Hilbert is proposing to constrain innovations by demanding a demonstration that systems expanded by the introduction of ideal, infinitary notions are conservative with respect to finitary mathematics. I think that this suggests an interesting, philosophically defensible position, which I shall now try to explain.

Section Five: Outline of a Hilbertian Philosophy of Mathematics. The position I am about to outline is hardly original. In particular, it bears a strong resemblance

---

[58] With some trepidation, let me offer the following (very partial) analogy. Think of Kronecker as a kind of radical physicalist, who interprets the unity of science as demanding full reducibility to elementary physics. You might then think of Hilbert as attempting to occupy a position like that defended in Fodor's well known article 'Special Sciences' (Fodor [1981]), in which token physicalism is combined with a denial that psychological laws can be reduced (even 'in principle') to purely physical laws - because (roughly) the law-like predicates of psychology do not subsume any physically unified class of objects. The analogy, I stress, is very partial, and intended to be no more than suggestive. If it does not help, then ignore it.

[59] The natural semantics for the ε-operator requires a domain of objects from which the choice function makes its selection. Since that domain might be, for example, the iterative hierarchy of sets, it is clear that this is can be a very powerful function - indeed, if ZF set theory without the axiom of choice is formulated with the ε-operator replacing the quantifiers in the background logic, a strong form of the axiom of choice becomes provable, *provided* the admissibility of ε-terms in the comprehension schema is allowed - a detailed treatment of this formulation of set theory is found in Bourbaki. (The need to modify the comprehension schema is very important. What the eliminability of ε-terms shows about the axiom of choice is that the real power of the axiom lies not so much in the permission of acts of arbitrary choice, but rather in permitting the assertion that the result of so choosing exists *as a set* - and for this, the modification of the comprehension schema is essential.)

to Quine's position in the philosophy of mathematics - unsurprisingly, since Quine and Hilbert are (I think) linked by a shared desire to accommodate mathematics within a broadly naturalistic philosophy, as well as by the shared influence of philosopher-scientists such as Mach and Herz, and their philosophical successors of the Vienna Circle.[60] Like Quine, but not quite for Quine's reasons, I shall argue that part, but not all, of mathematics ought to be understood literally. Here, however, we encounter one unpleasant obstacle in the path of progress towards the substantial philosophical issues, for when I say that part of mathematics should be understood literally, I am apt to be understood as having committed myself to 'realism' with respect to that part of mathematics. And that brings us to the term 'realism'.

Philosophical usage of this term is simply chaotic, and I do not propose to attempt to impose order on this chaos. Quine takes the acceptance of some parts of mathematics to entail commitment to mathematical objects, and thus he is a realist about those parts of mathematics in the sense of 'realist' (whatever it is) that appears to be in current use amongst philosophers of science. And he takes the acceptance of some other parts to bring with it no such commitments, and thus he is an anti-realist (in fact, a formalist) about those parts of mathematics, once again in the sense of 'anti-realist' that appears to be in current use amongst philosophers of science.[61] Now, if you think that Quine is a very queer kind of realist, rest assured that I agree with you. Quine thinks that the truth predicate is a device of disquotation, and that an explication of this fact constitutes an adequate philosophical account of truth.[62] If you think, as I do, that any worthwhile realism must involve some kind of correspondence theory of truth, then you will think, as I do, that Quine is in fact the arch anti-realist. Nevertheless, Quine is going to count as a realist about some parts of mathematics, in the sense in which I propose to use the term. When you read **Chapter Two**, you will see that my own realism (about part of mathematics) is more full-blooded than that of Quine, but for the moment let us ignore that. This is all that I propose to say about 'realism'.

I understand PRA realistically, in this Quinean sense. My reasons for so doing will be given largely in **Chapter Two**. But I also want to understand realistically all systems conservative over **PRA**, and this is the topic that concerns me at the moment. The

---

[60] I have neglected the important topic of the influence of Herz on Hilbert - for more on this, see Hallett [1990].

[61] See e.g. Quine [1984] p788.

[62] See e.g. Quine [1970] chapter 3, and especially Quine [1990].

questions that I shall now pursue, therefore, are these. Given a realist construal of PRA, what attitude should we take to the rest of mathematics? How far do our ontological commitments to mathematical objects go? Is there a principled way of extending a realist construal of mathematical systems beyond the finitary base provided by **PRA**, and if so, what is it? I take it that the attitude represented by (**H1**) does *not* constitute a principled way, and I shall give some (more) arguments in support of this view in **Section Six**. For the moment, though, I shall be looking for alternatives.

I have two reasons for approaching the issue of the ontological commitments of mathematics in this way, rather than by way of a detailed investigation into the semantic properties of mathematical language (in the manner of Dummett). To begin with, taking this approach enables us to stay reasonably close to Hilbert's own concerns and expressed philosophical opinions, whilst the alternative approach would force upon us very deep philosophical issues to which Hilbert was simply blind. Additionally, though, the arguments in favor of extensive versions of mathematical realism currently influential in the literature all stem from the Quinean 'indispensability for natural science' argument, and a large part of my purpose will be to show that that argument requires extensive supplementation by controversial doctrines in general metaphysics if it is to get you beyond the extremely restricted realism I endorse in this thesis.[63]

Let us begin by observing that Quine's opinions on ontological commitment to mathematical objects are no more mathematically revisionary than those of the out-and-out nominalist Field. What this fact immediately shows, I think, is that Hilbert's 'defence' of classical mathematics against intuitionistic and Fregean attacks persistently confuses two quite different kinds of issue. The first kind of issue concerns the *acceptability* of particular arguments and theories in mathematics. The second kind of issue raises the ontological question, the question of what acceptance of mathematical theories commits us to in the way of objects. The coherence of Quine's position shows that these two kinds of

---

[63] Of course, Quine's own views on the ontological commitments of mathematics themselves rest in part upon controversial views in general metaphysics. In taking this approach, I am therefore apt to be thought of as endorsing those views, and that is something I do not want to do. However, Quine's views on meaning, reference etc., although vastly more sophisticated than anything to be found in Hilbert, are nevertheless congenial to the concerns that motivate Hilbertian finitism, and therefore provide a relatively neutral basis for the present discussion. Furthermore - and this is the point to which I attach real importance - I think that attempts to stretch the Quinean argument for restricted realism to cover all of mathematics as it is standardly practised, in the manner of Penelope Maddy for example, will have to involve radical departures from Quine's basic metaphysic. I hope that my discussion will help to bring this out, and in this way show that a Maddy-style mathematical realist ought to be very interested in systems which respect the finitary conservativeness constraint basic to the Hilbertian approach.

issue are in fact orthogonal. Quine has (or need have) no complaints touching the mathematical *acceptability* of transfinite set theory, even at its outer reaches. But since he thinks that 'transfinite ramifications' are, in general, 'on a par with uninterpreted systems', acceptance of transfinite set theory does not commit him to, say, hyper-inaccessible cardinals. Hilbert - unsurprisingly - shows no awareness that issues of acceptability and ontology might pull apart in this way. Concerned to defend mathematics against an attack based upon an unwillingness to accept the apparent ontology of some classical mathematical theories, the Hilbertian who advances (**H1**) has presumed, wrongly, that the only, or at least the best way, to ensure acceptability is to defend the apparent ontology.

Once these two kinds of issue have been distinguished, however, it is possible to see (**H1**) in a somewhat better light. It is not altogether implausible, I think, to hold that a mathematical theory is *acceptable* if it meets the one structural constraint of consistency (perhaps relative to some theory we already have good reason to believe consistent) together with some broadly pragmatic constraints concerning such features as integrability within accepted parts of mathematics, fruitfulness of consequences, unifying and simplifying power etc. The verification that a theory meets these standards, notice, is indeed a purely internal affair, entirely governed by techniques proper to mathematics. It is quite another matter, though, to hold that our acceptance commits us to the apparent ontology of any mathematical theory which meets these constraints. The 'repugnance' that many philosophically inclined mathematicians feel towards (**H1**) (see e.g. Kreisel [1987] p395) is surely justified when (**H1**) is seen as a contribution to the solution of the ontological question.

What muddies the waters here, of course, is the issue of the appropriate *semantics* for mathematical theories. For as soon as we begin to press questions concerning, say, the relationship between acceptability and *truth*, or the appropriateness of literal *belief* in such and such a mathematical theory, the distinction between issues of acceptability and issues of ontological commitment is threatened with collapse. Suppose, for example, we establish the acceptability of some mathematical theory by formalizing it, showing that the formalization is consistent relative to set theory, and also has this, that and the other pragmatically desirable feature - by showing, in short, that (**H1**) is satisfied. Our acceptable theory with then prove theorems of the form $\exists x \varphi$. Have we now committed ourselves to objects (sets, as it might be) satisfying φ? Have we decided the issue of whether or not there are φ's? For any such theory, it will be possible to define (in a suitable metalanguage) a Tarskian materially adequate and formally correct truth predicate

for that theory. And if we accept (as we should) that a truth theory of this general kind ought to be the central component in a semantic theory for the object language mathematical theory, then are we not committed to the objects mentioned in the inductive clauses defining satisfaction by acceptance of the theory? If so, are we not obliged to say something about how the correctness of those clauses could ever be verified? If the answer to these questions is 'Yes', then the distinction between issues of acceptability and issues of ontology will indeed collapse.

This is a very familiar dialectic, and I shall have more to say about it below. For the moment, though, I will have to content myself with issuing a warning against a confusion induced by an ambiguity in the term 'interpretation' - a confusion which often induces a hasty positive answer to the above questions. In one sense, an interpretation of a language (or theory) is itself a mathematical object. It is a *function*, inductively defined on the primitive vocabulary of a given language L, taking values amongst the semantic values of the vocabulary of some language L' (which need not be distinct). Equally, the definition of a materially adequate and formally correct truth predicate is further piece of pure mathematics, this time resulting in the inductive definition of a predicate 'True (in L)' (in this case L and L' must be distinct).[64] Such definitions are contributions to semantics in the sense of *model theory*, but they are not in themselves contributions to semantics in any sense in which semantics involves study of the *import* a language has for the users of the language. Let us call this first kind of interpretation, *interpretation in the thin sense*. In a *different* sense of 'interpretation', however, an interpretation is a contribution to semantics in the sense of an attempt to state what knowledge of a language, and in particular knowledge of meaning, *consists* in. Call this, *interpretation in the thick sense*.

Now, ambiguities are rarely fortuitous. It would be surprising if there were to be no relation between the thick and thin senses of 'interpretation', and there is in fact good reason to believe that they are related in a way which is philosophically deep and important. Nevertheless, there is a dangerous and philosophically vicious confusion present in unguarded talk about 'interpreting' mathematical theories. If one interprets (in the thin sense) a mathematical theory T in set theory (say), one need not *thereby* abandon a purely formalist attitude towards T. Quine need have no complaints or qualms about the intensive studies currently being undertaken on models of set theories, and in accepting the results of such studies as parts of established mathematics, he certainly has not thereby committed

---

64 'True in the model M' is more commonly used than 'True in L'.

himself ontologically to the transfinite paraphernalia appealed to in the inductive specifications of the models. It is perfectly consistent for a formalist to be prepared to talk of interpretations of set theories in the thin sense, and it is perfectly possible for a strict formalist to think of this modelling activity as providing a notion of meaningfulness for mathematical theories. To think otherwise is simply to conflate the thin, purely mathematical and the thick, properly semantic senses of 'interpret'. Formalism, both as a matter of historical fact and as a matter of doctrinal content, takes issue with the interpretability of mathematics in the thick sense only.

However, Quine does not take a fully formalist attitude towards mathematics, and this is consequential upon his belief that some parts of mathematics do indeed have interpretations in the *latter*, thick sense. But the reasons why Quine takes this position have nothing to do with mathematics in particular: they have to do with Quine's opinions on the subject of meaning.

The immediate point is just this: In defending (**H1**) as a principle governing the interpretation of mathematical theories, one needs to be very clear about what one means by 'interpretation'. If you mean interpretation in the thin, mathematical sense, then something akin to (**H1**) might, in my view, be made plausible. But in defending (**H1**) in this sense, one will say nothing at all about any genuine ontological disputes. If you mean to speak of interpretation in the thick sense, though, then (**H1**) is indeed a contribution to ontological disputes, but a highly implausible one. I think that this should reinforce the contention that Hilbert, in much of his foundational work, is attempting something much more philosophically ambitious that what is strictly required for his defence of mathematics against the philosopher-revisionists. But of course, that is not to say that his attempts on the more ambitious goal are without interest, or without motivation. The motivation and the interest will quickly become apparent, I think, if we consider the reasons why Hilbert thought infinitistic mathematics required clarification and justification, for here we come across aspects of Hilbert's though to which philosophers have not always been sufficiently attentive.[65]

Hilbert [1925] makes much of the fact that progress in the natural sciences has left less and less room in the theory of nature for infinitistic notions, whether of the infinite divisibility of physical quantities or of the infinite extent of space (and time, although Hilbert does not

---

[65] The signal exception is Michael Hallett - see e.g. Hallett [1990].

mention this). Physical processes once thought to involve continuity essentially have been shown not to involve continuity at all, and as a consequence of this, ancient worries about the intelligibility of infinitistic notions to finite minds have been alleviated. And Hilbert thinks that a somewhat similar route has been traversed in mathematics itself - Hilbert [1925] begins, after all, with a handsome and extensive tribute to the work of Weierstrass, which he describes as establishing that the infinite as it appears in the notion of limit (as in the calculus) may be regarded as the potential infinite of the natural numbers series. The point here, in our terminology, is that, under analysis, a sentence such as (a)

(a) $f(x) \rightarrow \infty$ as $x \rightarrow 0$

can be seen not to require a reference for the apparent singular term ∞. However, Hilbert then points out that the price that has been paid for this achievement is the need for a theory in which the properties of *sets* of real numbers are studied. Since these sets may, and indeed typically do, involve infinitely many real numbers, such a theory will have to be considered as studying the properties of completed infinite totalities. And here, Hilbert suggests, the ancient worries reassert themselves, encouraged by the paradoxes of naive set theory.

But what ancient worries does Hilbert have in mind, exactly? The Gaussian claim that a priori cognition for finite minds must be limited to at most potentially infinite structures, perhaps - the claim that provides the grounding of (K2)? In detail, we do not know the answer to this, for Hilbert does not tell us in detail. However, there are a number of cryptic passages in his writings that give us something to go on.

Interspersed with the opening historical narrative in Hilbert [1925], there are hints of adherence to a Kantian thesis, according to which the antimonies of naive set theory, like those of naive dynamics or naive cosmology, are to be attributed to the illegitimate deployment outside the bounds of possible experience of principles applicable only within those bounds.[66] This can of course be seen as a concession to intuitionism: Hilbert is accepting that there is at least a prima facie doubt about the legitimacy of the law of excluded middle as applied to infinite totalities. Beyond this, though, it is very difficult to see what Hilbert intends by the Kantian analogy, since he never attempts to explain it at any length, or to give any account of his understanding of the Kantian principles upon which

[66] See Hilbert [1925] p376.

this claim about the legitimate deployment of concepts rests. And since the required explanations are hardly to be read off the agreed common consensus on the interpretation of Kant, we are still at a considerable disadvantage when it comes to trying to understand just what it was that Hilbert thought to be suspect about infinitistic notions, and why.[67]

However, there are also many passages in Hilbert's writings that make clear his desire for what we might call an *epistemologically credible account of the nature of our mathematical capacities*. And as Hilbert [1917] in particular makes abundantly clear, he thought that the axiomatic method had a crucial role to play in providing just such an understanding. In the striking opening passages of that article, for example, Hilbert emphasizes the importance, for the 'health' of any particular science, of preserving close integration with related fields of scientific activity. It is all the more striking that he illustrates this point with a claim that research in mathematics has been at its best when closely focussed upon problems and issues arising in the related fields of theoretical physics, and something he calls 'epistemology'.[68] Hilbert then goes on to describe axiomatization as a process in which the basic concepts of a science are isolated and their mutual interconnections displayed. Relations between closely related fields then emerge as relations between the concepts isolated in their respective axiomatic structures, occasionally revealing the possibility of a deeper unification within a single axiomatic structure.

Hilbert therefore seems to have thought that the 'health' of mathematics depended in part upon its relationship with a science he calls sometimes epistemology, sometimes psychology, sometimes logic.[69] And there seems to be little doubt as to his views about the nature of this relationship. Consider, for example, the following passage from Hilbert [1927]

---

[67] Let me emphasize once again that Hilbert really does have a concern with the underlying nature of paradox, with the etiology of paradox in the philosopher's sense, and not just with the problem of eradicating paradox from mathematical theories - see for example Hilbert [1917], where his evident satisfaction with Zermelo's 1908 axiomatization is not offered as sufficient grounds for dismissing worries about set theory.

[68] A central feature of Hilbert's scientific outlook is his denial of any principled distinction between pure and applied mathematics. He was quite passionately convinced that the imposition of such a distinction could only have ruinous consequences for mathematical research, and in his efforts to reform the German academic curriculum in accordance with this conviction, he found himself involved in an acrimonious and highly publicized controversy with the applied mathematician von Mises (see Reid [1986] pp342-343 e.g.). This in itself is enough to suggest that there is something very wrong with the received view of Hilbert as the arch formalist.

[69] In addition to Hilbert [1917] on this, see also e.g. Hilbert [1900] p3, Hilbert [1904] p131, Hilbert [1927] p475. Hallett [1990] gives some information on unpublished material from Hilbert's lecture notes on physics relevant to this topic.

> The formula game that Brouwer so deprecates has, besides its mathematical value, an important general philosophical significance. For this formula game is carried out according to certain 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 our 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 and phenomena deserves to be made the object of a serious and thorough investigation, it is this one .....[70]

Hilbert's claim is that the link between mathematics and this neighboring field of 'epistemology' occurs *at the level of finitary combinatorics*. In the finitary manipulations of symbols according to a few simple, precisely stated rules, the fundamental operations of the mind are expressed. Now, this is a very striking claim - bear in mind that this paper was written in 1927, twenty three years before Turing's paper 'Computing Machinery and Intelligence', long before von Neumann demonstrated the possibility of the programmable digital computer, and some forty years before the overthrow of behaviorism by cognitive science.

I suspect that this conviction concerning the 'technique of our thinking' may have prompted some such thought as the following. If the finitary manipulation of symbols is indeed the 'technique of our thinking', the transfinite content of the systems of classical mathematics becomes puzzling. There is no problem with characterizing our grasp of the potential infinite of the natural number sequence, for here the infinite is governed by finitary rules. But there does appear to be a problem about the actual infinite of set theory, where finitary minds are accredited with the capacity to grasp flatly infinitistic properties of complex objects - facts which appear to be independent of the rules according to which those objects are 'constructed'.

The appearance of a worry here, though, is surely the result of confusing the two senses of 'interpretation' mentioned above. We can do transfinite set theory, and nothing in these claims about the finitary 'technique of our thinking' is in any way inconsistent with this fact. When we do transfinite set theory - when, for example, we introduce an axiom postulating the existence of a cardinal number larger than any cardinal number constructible in a given axiom system - what we do is attend to the deductive consequences of the

---

[70] Hilbert [1927] p475. This passage is the most striking expression of Hilbert's opinions on the 'general philosophical significance' of axiomatization and the metamathematics study of axiom systems, but there are strong suggestions of these views in his writings at least as early as Hilbert [1900].

axioms, as well as such matters as their simplifying power, fruitfulness for mathematical theory in general etc. This, however, is all at the level of acceptability of mathematical theory. Nothing in our ability to do transfinite set theory demands the introduction into our ontology of large cardinals, and nothing in this ability suggests some kind of mysterious direct insight into the properties of infinitely complex objects. Nothing in the acceptability of set theories with or without very large cardinal axioms of itself demands any interpretation of the set-theoretic axioms in the strong, properly semantic sense. If Hilbert thought that we must either abandon set theory as we now have it, or accept (**H1**), then I think that he was simply mistaken. If he thought that acceptance of set theory brought with it puzzles about the epistemic access of finite intelligences to infinitely complex objects, then I think that he was confused.

However, whilst I do think that these mistakes and confusions are present in Hilbert's thought, I also think that his convictions about the 'technique of our thinking' and its implications for our understanding of the nature of mathematics also prompted some important insights. For there is still the Quinean argument in favour of genuine ontological commitment to at least some mathematical objects, and the problem of giving an epistemologically credible account of our epistemic access to those objects is, I think, a genuine one. For the fact of the matter is that the only remotely credible scientific theories of cognition we have are indeed of the kind Hilbert foresaw. Insofar as the mind has proved amenable to empirical scientific study at all, it has proved to be an information-processing device, not essentially different in character from a digital computer. Now of course, it would be much too quick to conclude from this that there was so much as a prima facie case for a philosophical obligation to explicate our cognitive capacities entirely in information-processing terms. For the information-processing accounts of our cognitive capacities given to us by empirical science concern themselves with our empirical knowledge, and it is not unreasonable to think that our deductive capacities demand a somewhat different approach. But now the problem of ontological commitment to mathematical objects becomes serious, for (contra (**H1**)) our knowledge of mathematics cannot be assimilated within an account of the workings of the deductive capacity in general. The problem, of course, is the proprietary ontology of those mathematical theories to which the Quinean argument commits us: unlike logic, mathematics is not topic neutral.

What is this Quinean argument? In essence, it is this. Ontology is subordinate to empirical science, and not vice versa. Our best - indeed, our only - guide to what there is is what empirical science tells us there is. But empirical science commits us to at least some

mathematical objects, since the literal truth of physics (for example) as we now have it requires mathematical objects - real numbers, for example - in addition to physical objects as the values of bound variables. (In the Putnam variant of the argument, it is pointed out that we cannot so much as *state* the laws of physics without committing ourselves to real numbers.[71])

But why should it be *empirical* science that enjoys this privileged position with respect to ontology? Why not science in general - in which case commitment to all the objects quantified over in accepted mathematics immediately follows? A detailed answer to this question would take us deep into Quine's theory of meaning, and I cannot attempt to give such an answer here. But the crucial component of the Quinean answer is the need for the scientific study of issues concerning meaning and reference *to be adequately grounded in the intersubjectively available evidence of the senses.*

In the famous Quinean image, the 'lore of our fathers' comes down to us in an articulated web of sentences, assent to those at the periphery being closely conditioned to occurrent stimulations of the sensory receptors, whilst assent to those closer to the center becomes progressively more robust under perturbations of sensory input.[72] Together with this image comes a just-so story of how languages might be learned, in which words for ostensible objects and properties are gradually supplemented with ever more recherché items of vocabulary in a spontaneous, ever-revisable outburst of theory designed to render the ongoing flux of sensation cognitively manageable.[73] This process of spontaneous theorizing associated with what Quine calls the 'positing' of physical objects over and above the 'phenomenal' objects immediately present in sensory stimulation

> Physical objects are conceptually imported into the situation as convenient intermediaries - not by definition in terms of experience, but simply as irreducible posits comparable, epistemologically, to the gods of Homer. For my part I do, qua lay physicist, believe in physical objects and not in Homer's gods . . . But in point of epistemological footing the physical objects and the gods differ only in degree and not in kind. . . . Positing does not stop with macroscopic physical objects. Objects at the atomic level are posited to make the laws of macroscopic objects, and ultimately the laws of experience, simpler and more manageable . . . . Physical objects, large and small, are not the only posits. forces are another example; and indeed we are told nowadays that the

---

71 See Putnam [1979] p337-341 e.g.

72 The full story is given in Quine [1960], especially in chapters 1 - 3.

73 For the just-so story in detail, see Quine [1974].

> boundary between matter and energy is obsolete. Moreover, the abstract entities which are the substance of mathematics . . . are posits in the same spirit. [74]

Now this Quinean story becomes controversial at the level of detail, but I think that the broad outlines are largely independent of the doctrines associated with translation into canonical idiom. And at least at the level of this broad outline, the Quinean story would surely be congenial to Hilbert. For the guiding intention to integrate an account of mathematics and reference to mathematical objects within an overall understanding of the place of our cognitive capacities within the natural scheme of things, and above all the desire for a close integration of the theoretical concepts of physics and mathematics, are deeply Hilbertian in spirit - as is the desire to free scientific theorizing from dependence upon any kind of transcendental philosophical support.

There are, however, two respects in which the Quinean story seems counterintuitive. In the first place, it appears to have the consequence that the existence or non-existence of mathematical objects of such and such a kind becomes an empirical matter, to be determined ultimately by the development of theoretical physics. But it seems strange that physics should serve as the guarantor of the existence of such objects, and yet be unable to tell us anything of their properties.[75] And in the second place (as Charles Parsons has repeatedly pointed out) the Quinean picture appears to locate all of mathematics within the most deeply theoretical parts of total science, alongside the extremely recondite truths of fundamental physics. And that seems to render mysterious the *obviousness* of at least the most elementary parts of mathematics.[76]

---

[74] Quine [1951], p44-45. A more careful telling of the Quinean story would remove any hints of phenomenalism.

[75] Penelope Maddy, in Maddy [1991], complains that Quine's account leaves the justification of some mathematical truths in the hands of physicists (see p31). But it does not: Maddy has ignored Quine's views about the correct philosophical account of truth, and thus falls into the trap of conflating issues of scientific acceptability and issues of ontological commitment. As we have seen, these are orthogonal for Quine. This is not my complaint. She also claims [loc cit] that Quine's position leaves unapplied mathematics without justification: and this, of course, is a further consequence of the same conflation.

[76] See Parsons [1983] e.g. I do not mean to imply that Quine has no response to this complaint. The Quinean metric of entrenchment is normally taken to articulate the understanding of an ordinary member of the speech community, and with respect to such a citizen, the obviousness of the sentences expressing the truths of elementary mathematics, like the obviousness of all simple tautologies, is reflected in the fact that those sentences are highly robust under peturbations of sensory input. The sentences which express observational facts are not in this way robust, and of course such sentences are not in the same sense obvious. Prompting me right now with the sentence 'This apple is red' will indeed produce ready assent, since it is indeed obvious to me that this apple is red. But no stable overall pattern of assent to that same sentence (type) can be generated, since (for example) it is by no means obvious to you, right now, that this same apple is red - after all, you can't see it. Lila Luce, in a recent article devoted to this alleged problem of the obviousness of elementary mathematics, seems to me to be confused on these matters - see Luce [1990].

Now, the indispensability for science - or more exactly, for all 'scientific thought' - of *finitary* mathematics is in fact a Hilbertian theme. In **Chapter Two**, I shall try to explain what Hilbert may have had in mind by this indispensability thesis, and argue that it is very plausible. The interest of the Quinean view then lies in the suggestion it contains as to a principled, broadly naturalistic way of extending a literal construal of finitary mathematics to a literal construal of at least some parts of infinitary mathematics. The suggestion is that the same kind of process - roughly, general scientific rationality - that generates ever more sophisticated physical theories from the evidence of the senses, and thus introduces into our ontology ever more recondite species of physical objects without violating the underlying constraints imposed by a naturalistic construal of our cognitive capacities, will also permit the introduction into our ontology of more recondite species of mathematical objects within those same naturalistic constraints.

In the Quinean story, the introduction of objects is a theoretical speculation, constrained by experimentation - the evidence of the senses. In mathematics, however, there are no experiments. What, then, is the relevant constraint? One answer is given by (**H1**): the sole relevant constraint is consistency. But that is not naturalistically plausible, since it is not plausible at all. In Hilbert's later writings in particular, though, there is the suggestion of a better answer. Consider this passage, in which he responds to the Brouwerian complaint that metamathematics is 'empty formalism':

> This formula game [i.e. deduction in a formalized theory] enables us to express the entire thought-content of the science of mathematics in a uniform manner and develop it in such a way that, at the same time, the interconnections between the individual propositions and facts become clear. To make it a universal requirement that each individual formula then be interpretable by itself is by no means reasonable; on the contrary, a theory by its very nature is such that we on 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. Only certain combinations and consequences of the physical laws can be checked by experiment - just as in my proof theory only the real propositions are directly capable of verification.*[77]

For the representative citizen, of course, the truths of theoretical physics are not in this sense robust either, since attempts to elicit response to sentences of theoretical physics will result in bizarreness reactions. The reason that I introduce the obviousness complaint against Quine is not that I think that his position is powerless to respond, but rather that I think that the response Quine can make will end up drawing heavily upon the more controversial aspects of Quinean doctrine. Since the position that I am currently labouring to outline involves a rejection of those more controversial aspects, I think it worthwhile to point out this prima facie problematic feature of Quine's views.

[77] Hilbert [loc cit], my emphasis.

The suggested analogue of experimental evidence, then, is *computation*. The 'direct' verification of a mathematical proposition proceeds by computation. The 'real' propositions - the propositions of finitary mathematics - are those which can, at least in principle, be verified by computation, in the same way that the experimental consequences of a physical law can be verified (in principle) by observation. The 'ideal' elements, in so far as they can be eliminated from proofs of finitary theorems, may be regarded as theoretical terms, in the sense that they function essentially as unifying and simplifying agents, without exceeding the boundaries of the computable. Thus constrained, the mathematical theory cannot take us beyond what we could in principle do by non-theoretical means.

Notice that the claim here is not that ideal elements may be regarded as theoretical terms in the sense that their meaning must be explained holistically in terms of the role they play in an axiomatized system. Given Hilbert's views about the axiomatic method, the concepts of finitary mathematics are every bit as theoretical in that sense. There is no giving an account of the concept of *natural number* independently of notions such as *successor* which feature in the arithmetic axioms: arithmetic has no observational concepts. Rather, the suggestion here is that our ability to interpret mathematical theories in the thick sense is constrained, not only by the deductive consequences of arbitrarily postulated axioms in the manner suggested by (**H1**), but also by the bounds of what can in principle be done by finitary resources. Within those bounds, we are free to investigate the ideal superstructure with all available means. Beyond those limits, though, there is no reason to think of mathematical theories as demanding interpretation in anything other than the thin sense.

Now in the Quinean picture, mathematics stakes its ontological claims in the process of our attempts to explain and predict the ongoing flux of experience, in the same process that produces the theoretical entities of physics. The parallel within mathematics of this process would be one on which the actual infinite stakes its ontological claims in the process of our attempts to explain the properties of finitary mathematics - to smooth out and simplify the theory of the 'observational' data available in principle by computation. The resultant picture of mathematics would be one on which we incur commitment to the objects of finitary mathematics in some direct way, and commitment to infinitary objects insofar as they may be regarded as simplifying devices. This is a restricted form of realism, as we have been using that term - one in which commitment to the actual infinite is restricted to those infinitistic theories that are conservative over finitary mathematics. It is a realism that

promises to encompass all the mathematics that is required for our theory of the physical world, but it commits us to ideal mathematical objects in response to a mathematical need. There is no puzzle as to how physics can provide the raison d'être of these objects without being able to tell us anything of their properties, for physics does not provide the raison d'être of those objects.

But why should we respect the restriction on our ontological commitments imposed by the conservation constraint? Why not accept commitment to all of transfinite set theory in the same spirit? After all, set theory arose precisely out of an attempt to deepen and generalize our grasp of the properties of the real number system, at least some of which we are already committed to on the approach that accepts the conservation constraint. Why does this not provide sufficient reason for a full-blown set theoretic realism?

This question has a close analogue in the philosophy of the natural sciences. It has been a persistent criticism of the constructive realism advocated by Bas van Fraassen, for example, that it allows the actual observational powers of human beings to play an inflated role with respect to our ontological commitments in natural science.[78] For van Fraassen, acceptance of a mature theory in the natural sciences commits us only to those parts of the prima facie ontology of the theory which are observable in principle. Beyond this, acceptance of the theory as empirically adequate allows us to remain noncommittal with respect to the existence of unobservables. Acceptance of atomic theory as empirically adequate commits us to molecules, perhaps; but not to photons. I shall have more to say on this point towards the end of the next section.

Section Six: Clarification and Comparisons. This sketch of a position on the ontological commitments of mathematics, and the notion of meaningfulness required for mathematical theories, certainly demands extensive clarification and defence. Some of that defence will have to be postponed for the final section of **Chapter Two**, for it involves crucially the special status the Hilbertian associates with finitary mathematics - the topic of **Chapter Two**. I want to finish this chapter, though, by offering some preliminary clarification and defence, mostly in the form of a comparison with the views of other authors, both on Hilbert and on mathematics in general. For the most part, I shall be arguing against apparently richer form of mathematical realism accepted by Maddy, and the

---

[78] Van Fraassen's views are set out and defended in detail in van Fraassen [1980]. For discussion of van Fraassen's views on unobservables in particular, see the essays collected in Hooker and Churchland [1985].

formalism of Tate. But I shall also have something to say on the apparently thinner understanding of Hilbertian philosophy of mathematics offered by Detlefsen in his book on Hilbert's Programme.

In Tait [1986], one finds a good contemporary statement of a philosophy of mathematics which adheres to something very like (H1) as a sufficient response to questions concerning the ontological commitments of mathematics. For Tait, the thin notion of interpretation provides a philosophically adequate account of mathematical practice. He writes:

> . . . Platonism does not consist in an interpretation of mathematical theories. We do indeed interpret theories in mathematics, as when we construct inner models of geometries or set theory or when we construct examples of groups etc., with certain properties. But we do this in the language of mathematics, and our 'grasp' of this consists in our ability to use it . . .. Benacerraf and Putnam seem to me to be typical of those who adopt a particular picture of Platonism. The picture seems to be that mathematical practice takes place in an object language. But this practice needs to be explained. In other words, the object language needs to be interpreted. The Platonist's way to interpret it is by Tarski's truth definition which interprets it as being about a model - a Model-in-the-Sky - which somehow exists independently of our mathematical practice and serves to adjudicate its correctness. So there are two layers of mathematics: the layer of ordinary mathematical practice in which we prove propositions such as ['there is a real number greater than 10'] and the layer of the Model at which ['there is a real number greater than 10'] asserts the 'real' existence' of a number. . . . [This is not the] version of Platonism that I am defending or that I even understand. . . . Tarski's truth definition . . . is a piece of mathematics, concerning the mathematical notion of a model of a formal language.[79]

As an account of the acceptance conditions for mathematical theories, I have great sympathy with Tait. But I differ from him in that I do not believe that the thin notion of interpretation invoked here provides a philosophically adequate account of the ontological commitments of mathematics. I have two reasons for this.

The first concerns the applicability of mathematics - the feature of mathematics that is central to the Quinean indispensability arguments. The basic point here is familiar, since it has been the central complaint against formalism since Frege. Tait, I maintain, can give no adequate account of the applicability of mathematics in natural science. Tait considers the issue of the applicability of mathematics, and offers the following account:

> Consider . . . a mathematical prediction of the motion of a physical object. First, we read the appropriate equations off the data - i.e. we chose the appropriate idealization of the phenomenon. Second, we solve the equations.

---

[79] Tait [1986] pp348-349.

> Third, we interpret the solution empirically .... [Mathematical knowledge here] is knowledge that S, where S is a mathematical proposition. But *that* kind of knowledge is involved only at the second step, and it involves nothing empirical. The first and third steps involve only knowing how to apply mathematics to the phenomena.[80]

The kind of use of mathematics Tait speaks of here is typical of theoretical science, in which one constructs a mathematical model of some natural system, and derives predictions about future states of the system (for example) by purely mathematical reasoning in the model. Here, it is possible to separate out the mathematical from the empirical components in the scientific theory. But that is not the only way that mathematics gets applied in science. At the very simplest level, any practical application of science will involve the measurement of some physical quantity. For example, the engineer's explanation of why the bridge collapsed under the strain of carrying fifty trucks will involve a number of claims which, if transcribed into canonical notation in Quine's fashion, will involve predicates, the satisfaction conditions of which require sequences composed out of a mixture of physical objects and real numbers. Here, the mathematical component of the explanation cannot be hived off from the physical component. They are inextricably mixed in the engineer's reasoning. It seems to me that the correctness of the engineer's explanation requires an interpretation of his utterances in the thick sense, and that, in turn, induces genuine ontological commitment to real numbers.

My second complaint against the Tait account of the nature of mathematics is that it seems to me to make the applicability of mathematics something rather surprising, something one would not be led to expect from any examination of the internal practice of mathematics. For we are given the impression that mathematics is an essentially self-contained activity. Tait seems to have no room to acknowledge that the great concentration of mathematical activity on systems which do have application in the physical world has anything other than a purely extrinsic explanation, perhaps in terms of quasi-institutional pressures, or psychological facts to do with the mathematical imagination. And that in turn makes it a kind of fluke, a cosmic coincidence, that the universe obeys mathematical laws. Lucky old us. And that is surely wrong. I do not believe that a mathematically lawless universe, and a fortiori *experience* in a mathematically lawless universe, is a genuinely intelligible possibility.

---

[80] Tait [op cit] p351.

Whatever its shortcomings, the Hilbertian view I have been sketching does not have these problems. On the Hilbertian account, mathematics is intrinsically connected to experience and to the physical world. The applicability of mathematics is taken as a constitutive feature of mathematical activity, and it is central to our account of the understanding of mathematical theories. Finally, providing all of applicable mathematics can indeed be formalized in systems conservative over **PRA**, as has not been established, the requirements for a thick interpretation of applicable mathematics have been in some measure addressed, without the invocation of a Model-in-the-sky.

Tait serves as an example of a philosopher who wishes to limit the demand for interpretation of mathematical theories to interpretations in the thin sense. Penelope Maddy, on the other hand, appears to demand a thicker notion of interpretation for almost all of mathematics as it is currently practised, including perhaps some very powerful extension of standard **ZF**. In Maddy [1990], she draws on some well-known suggestions made by Gödel to outline ways in which a mathematical realism grounded in the Quinean indispensability argument might be extended far beyond the most generous boundaries countenanced by Quine, deep into the regions populated by cardinals larger than any whose existence can be proved in **ZF**.[81]

Maddy, however, seems to me to be occupying a highly unstable position, for I think that she does not realize the extent of her departure from the theses that provide the fundamental underpinnings of the Quinean indispensability argument. Given that her basic argument for realism depends entirely on that argument, I think that she must either give up her demand for a more extensive realism and keep faith with the Quinean indispensability argument, or break with Quine altogether, in which case her position seems to me likely to collapse into that of Tait. Let me explain.

Maddy insists that her position is based upon a Quinean naturalized epistemology: we are to do philosophy standing 'within our own best theory of the world', and epistemology accordingly becomes a 'descriptive and explanatory' project.[82] We have already seen how this perspective is associated with the Quinean indispensability argument - a modicum of mathematics is involved in our best natural scientific theory, and from the perspective of

---

[81] The Gödel suggestions come by and large from Gödel [1964], and are contained in a passage which I shall discuss extensively in the third section of **Chapter Two**. Maddy's views on these matters are explained at greater length in her two part paper Maddy [1988a] and [1988b].

[82] Maddy [1990] p9.

epistemology naturalized, the arguments that commit us to the physical objects quantified over in that theory also commit us to the mathematical objects.

Given that this is the perspective, though, it is very curious to find Maddy complaining, against Quine, that unapplied mathematics 'is completely without justification on the Quine/Putnam model; it plays no role in our best theory, so it need not be accepted'.[83] Quine's indispensability argument is not intended to *justify* any part of scientific theory. It is not intended as a measure of the *acceptability* of any scientific theory, whether natural or formal. How could it possibly be that, compatibly with the perspective of epistemology naturalized? The *justification* of mathematical theories is for mathematicians, not philosophers. The *acceptability* or otherwise of a mathematical theory is a matter to be determined by the internal practice of mathematics. The Quinean question is this: Given the perspective of epistemology naturalized, what is the minimum ontological commitment demanded of us by natural science? And for Quine, that minimum commitment is measured by a procedure that translates our 'theory of the world' - natural science - into canonical idiom, and then looks at the domain of the quantifiers. Canonical idiom here is first order quantification theory with identity, with all quantifiers understood objectually. For any theory not in canonical idiom, or indeed for any theory which has no role to play in our best theory of the world, the ontological question, for Quine, has no determinate answer.[84] What is more, the translation procedure into canonical idiom is subject to the indeterminacy of radical translation. At the level of basic physics, leave alone that of mathematical theory, that indeterminacy is extensive.

The fact of the matter now is that very little mathematics appears as a canonical commitment of physics. Quine regards himself as committed to sets, since he regards all mathematical objects as sets (recall the indeterminacy of radical translation if you are inclined to take an interest in whether real numbers are 'really' sets). But the set theory one is committed to by physics in canonical idiom is minimal: the exact measure of that commitment I do not know, but it is certainly to a vanishingly small fragment of **ZF**. Beyond this minimum, issues of the ontological commitments of mathematics, as I understand Quine, really have no determinate answer. This does not mean that the parts of mathematics - including almost all transfinite set theory - that never appear in the translation of our theory of the world into canonical idiom is in any way defective as mathematics, or in need of some

---

[83] Maddy [op cit] p30.
[84] See Quine [1969] p106, e.g.

justification. It is just that, as Quine understands matters of ontology, in accepting those parts of mathematics, you incur no ontological commitments. Transfinite set theory, for the most part, is 'on a par with uninterpreted systems'.[85]

The confusion we have been unearthing here is the confusion we encountered above, between issues of justification and issues of ontology. On the Quinean picture, these issues are orthogonal, and this is implicit in the perspective of epistemology naturalized as Quine presents it. Our ontological commitments are to be determined from within our best theory of nature. But then, how is Maddy to get genuine ontological commitment in transfinite set theory from the Quinean indispensability argument?

In her work on the status of the axioms of set theory, including axioms beyond those of standard **ZF**, Maddy offers an extensive and fascinating discussion of a procedure she calls 'extrinsic' justification of axioms - justification in terms of their unifying power, their fruitfulness, there simplificatory potential within mathematical theory. She points to the undoubted distortions induced in standard analysis if use of the axioms of dependent choices (at least) is abjured:

> In the case of the axiom of choice, then, we have our first example of an extrinsic defence of a set theoretic hypothesis, beginning with a straightforward indispensability argument: our best theory of the world requires arithmetic and analysis, and our best theory of arithmetic and analysis requires set theory with at least the axiom of dependent choice.[86]

She describes this as 'pure Quine/Putnamism', but it is nothing of the kind. There is no general argument in Quine (or the time-slice of Putnam to which Maddy refers) that licenses an inference from best scientific theory to ontological commitment. There is an argument from best theory of nature to ontological commitment, but that is a very different thing. There are three points to note here. In the first place, given a mathematically elegant and fruitful formalization of the mathematics needed in the theory of nature which is profligate with sets, and a mathematically clumsy and barren theory which is economical with sets, there is nothing in Quine's position that need incline him to prefer the former. Secondly, though, and much more important, the indeterminacy of radical translation should, in these circumstances, incline the Quinean to view with great suspicion any claim that there is a genuine ontological issue to decide. After all, Quine thinks that we are committed to some

---

[85] See Quine [1934] p788.
[86] Maddy [op cit] p120.

set theory by the theory of nature, but there is no fact of the matter as to *which* set theory we are committed to. Your choice between ML, NF, ZF, NBG, Kripke-Platek, Kelley-Morse or whatever may surely be made on grounds of mathematical convenience - or, if you will, fruitfulness, simplicity etc. - but such grounds, for Quine, equally certainly cannot induce any inflation of ontology. And thirdly, it is a mistake to think that Quine is prepared to countenance an inference from best scientific theory to ontological commitment even in the case of *natural* science. For ontological commitment, I repeat, is determinate only for theories in canonical idiom, and in at least two very well known areas of natural science as we currently have it, Quine would repudiate any suggestions of ontological commitment in virtue of the fact that an acceptable translation into canonical idiom is impossible. I speak of cognitive psychology and linguistics.

This last example is instructive, for it speaks to any attempt to broaden the Quinean indispensability argument to one on which we are ontologically committed to the entities quantified over in currently accepted scientific theories in general. It happens that I would be rather sympathetic to this view, but it faces at least two difficulties. The smaller concerns determining the appropriate notion of canonical idiom, for some such notion is needed if the proposal is to yield any determinate result. Much more important, though, it is very doubtful if this can be accommodated to the perspective of epistemology naturalized as Quine understands it. The difficulty, of course, lies in the fact that this inflation of the indispensability argument threatens the Quinean with a resurrection of the analytic/synthetic distinction in some form. This is in fact vivid in Maddy's discussion of the axioms of set theory, for it is very hard to avoid the impression that she thinks certain axioms can be justified by demonstrating that they keep faith with the mathematical concept of set. I am, I repeat, very sympathetic to this line of thought, but it does not seem to me to be compatible with naturalized epistemology as Quine understands it, since it invokes a notion of analytic entailment which Quine would, I think, would regard with suspicion. And if such a line of thought is used to invoke commitments to objects, then the break with the Quinean perspective is definitive.

If Maddy is to maintain this line of thought on the axioms of set theory, then, she cannot rest her fundamental case for commitment to mathematical objects on the Quinean indispensability argument. Her account of the ontological commitments of mathematics will then of course depend upon which alternative foundational argument she chooses, but if the result is to be the set theoretic realism she currently avows, then I think that she will face one of two dangers. The first is a Tait-like formalism. The second, though, which is

strongly suggested in her discussion of the axioms of set theory, is the kind of Fregean dogmatism that Hilbert so strongly opposed.

Consider, for example, Gödel's axiom of constructibility. This axiom restricts the iterative generation of sets at each stage to those explicitly definable by predicative formulas. The result is a fragment of the full standard ZF universe, in which the axioms of ZF without choice all hold, and choice is provable. The resulting theory has very nice properties. As Maddy says, all the outstanding issues in analysis, and even such controversial matters as the generalized continuum hypothesis are provided with determinate answers. However, it also has some 'counterintuitive' consequences, which incline many set theorists to work in a richer universe, in which the combinatorial generation of sets is not restricted by predicative considerations. Is there an issue to be adjudicated here? There is, of course, an internal issue, concerning the relative fruitfulness, simplicity, etc. etc. of these two systems. But is there an issue which calls for adjudication in some different sense - a decision as to what the set-theoretic universe is 'really' like? Maddy appears to believe that there may well be. Having described some other additions to and departures from standard ZF, she writes:

> I have described two theories, two extensions of ZFC, that cannot both be true. Each theory answers at least the open question of Luzin and Suslin, and one even decides the size of the continuum. Each enjoys an array of extrinsic supports . . .. The philosophical open question is: on what rational grounds can one choose between these two theories.[87]

That may be *a* philosophical open question, but it surely is not *the* philosophical open question. One thing one might also want to know is, why is a choice between these theories thought to be necessary at all? Of course, if you think that at most one of these theories can be true, then you do indeed have to face Maddy's question. But why should one think that? The echoes of Frege on non-Euclidean geometries are very striking here. And the Hilbertian reply is the same in both cases. Given a notion of truth tailored to the thin notion of interpretation, there is no reason to believe that at most one of these theories is true, since truth here is no more than truth in a given model, and there is no reason why there should not be many models. And if you insist on the thicker notion of interpretation, then you will end up (with Frege) leaving mathematics incapable of deciding matters that clearly belong within its purvue, facing perplexities that cannot be solved because they

[87] Maddy [op cit] p143.

cannot be stated in a vocabulary the correct interpretation of which can be stated to the mutual satisfaction of the mathematical community.[88]

The Hilbertian position I have been sketching does not have any of these difficulties. The fundamental argument for the existence of mathematical objects, for the Hilbertian, is not the Quinean indispensability argument, and the underlying perspective is not that of epistemology naturalized as Quine understands it. And Fregean dogmatism is avoided by repudiating any suggestion that mathematics, beyond the bounds of the potentially calculable, has any need of interpretation in the thick sense.

But here the advocate of a thicker realism may see a serious weakness in the Hilbertian position. After all, set theory arose precisely out of an attempt to deepen and generalize our grasp of the properties of the real number system, at least some of which we are already committed to on the approach that accepts the conservation constraint. Why does this not provide sufficient reason for a full-blown set theoretic realism? Why concern ourselves with the bounds of the calculable in principle?

This is the question raised at the end of the last section, and as I observed, it has an interesting analogue in the philosophy of natural science. It has been a persistent criticism of the constructive realism advocated by Bas van Fraassen, for example, that it allows the actual observational powers of human beings to play an inflated role with respect to measuring our ontological commitments in natural science.[89] For van Fraassen, acceptance of a mature theory in the natural sciences commits us only to those parts of the prima facie ontology of the theory which are observable in principle. Beyond this, acceptance of the theory as empirically adequate allows us to remain noncommittal with respect to the existence of unobservables. Acceptance of atomic theory as empirically adequate commits us to molecules, perhaps; but not to photons.

But why, van Fraassen's critics persist, should we allow the evidently local and changing boundaries of the perceptible-by-us be allowed to play so important a role in determining what there is? Indeed, why should those boundaries be allowed to play any role at all? Now, whilst the analogy with the mathematical case, in which the calculable-in-principle

---

[88] I must confess that I find it genuinely extraordinary that Maddy, in a book which treats Quinean naturalism with such deference, can so much as suggest that there might be a serious issue as to which set theory is true.

[89] See e.g. Gutting, G. 'Scientific Realism versus Constructive Empiricism: a Dialogue', in Hooker and Churchland (eds) [1985] pp118-131, and van Fraassen's reply ([op cit] pp252-258.

takes the place of the observable-in-principle, is less than perfect (since the former has more of the appearance of an absolute distinction than the latter has) there is nevertheless enough of an analogy, to allow us to profit from van Fraassen's response to this complaint. He writes:

> If I believe [a scientific theory] to be true and not just empirically adequate, my risk of being shown wrong is exactly the risk that the weaker, entailed belief [i.e. the belief that the theory is empirically adequate] will conflict with actual experience. Meanwhile, by avowing the stronger belief, I place myself in the position of being able to answer more questions, of having a richer, fuller picture of the world, a wealth of opinion so to say, that I can dole out to those who wonder. But, since the extra opinion is not additionally vulnerable, the risk is - in human terms - illusory, and *therefore so is the wealth*. It is but empty strutting and posturing, this display of courage not under fire and avowal of additional resources that cannot feel the pinch of misfortune any earlier. [90]

Under the terms of our analogy, we may say the point being made here is that, beyond a certain stage, claims about the availability of interpretations in the thick sense for mathematical theories become philosophically idle. Despite the realist's thumping on the table, no further theoretical commitments are in fact being made. This point is not a verificationist one, notice: the claim is not that talk of thick interpretation beyond the critical stage becomes meaningless. Axioms postulating the existence of this, that, or the other kind of large cardinal need not be accounted meaningless even by a Quinean, for the Quinean can hold that interpretation in the thin sense provides a perfectly adequate notion of meaning for mathematical statements, and such axioms can, in this sense, be provided with interpretations.[91] The point is that the apparent extra risk of richer forms of realism, the air of affirming a controversial philosophical position by insisting on interpretation in the thick sense, is spurious. The only constraint that such talk is in risk of violating is that imposed by (H1), and mathematical realism is normally thought to involve the satisfaction of some much stronger condition.

Let me try and clarify this a little.[92] Consider, for example, the claim that there are planets outside the light cone of the human race, and let us suppose that this is an implication of cosmology. On a van Fraassen type position, as I understand it, your acceptance of that cosmological theory does not commit you to those planets - you should remain agnostic with respect to their existence. And this is because, in van Fraassen's terms, the apparent

---

[90] Van Fraassen, in Churchland and Hooker [1985] p255.

[91] Of course, this cannot be empirical meaning, so it is powerless to have any effect on ontological questions.

[92] I am indebted to Martin Davies here for helpful discussion.

extra degree of risk involved in ontological commitment to those planets is illusory. Appearances to the contrary notwithstanding, the realist's risk factor is in fact just the same as the van Fraasenite's. Now, this talk of risk factors undoubtedly suggests verificationism. However, think of the matter this way. The sentence expressing the realist commitment is perfectly meaningful, since it is composed out of meaningful components in a meaningful way. In the language of risk factors, the component concepts of the claim are not themselves at zero risk: zero risk is in this case the joint product of the laws of physics and non-zero risk conceptual components. Because of this, the realist air of bold commitment is merely apparent.

In the same way, then, I say that the realist's air of bold commitment to thick interpretation of extensions of set theory dealing with, say, hyper-inacessable cardinals is merely apparent. The extra commitment is associated with no greater genuine mathematical wealth. The mathematics involved is perfectly meaningful, and may be as rich and rewarding as you please. However rich and rewarding it is, though, it becomes no jot more rich and rewarding for the insistence on the availability of a thick interpretation. The air of holding mathematics hostage to the satisfaction of some stronger constraint than the Hilbertian will grant is bogus.

Now, if something like this van Fraassen-type position (which I take to be at least close to Quine's own position) is plausible with respect to natural science, it must surely be the more plausible with respect to mathematics. Observation, after all, is a causal notion, and the limits of observability are therefore fixed by the same scientific laws that predict the existence of matter outside the light-cone of the human race. But such as it is, this source of hope for a richer form of scientific realism is surely closed off to mathematical realism, since there are no plausible suggestion as to what might count as an genuine analogue of causation to provide some extra determinacy in the outer reaches of the set theoretic universe. Maddy, for example, thinks that we can causally interact with some very small sets, and perhaps learn the basic principles of set theory via this causal interaction. But even she allows that the exotica of transfinite set theory can in principle only be reached by deep mathematical theory. My contention is that she has provided no support for the contention that the crucial mathematical generalization into the transfinite, beyond the bounds of the calculable in principle, can commit us ontologically to the objects of transfinite set theory.

I wish to conclude now with a few words on Detlefsen's allegedly Hilbertian philosophy of mathematics, and in particular on the subject of instrumentalism. Detlefsen advocates instrumentalism, and thinks that Hilbert was himself an instrumentalist. I do not agree with the latter point, and I do not understand the instrumentalism that Detlefsen advocates. Now, my main discussion of Detlefsen will be given in **Chapter Three**, because his attack on the accepted interpretation of the incompleteness theorems is crucial to his defence of his position. In these closing paragraphs, then, I shall in part be setting the stage for later discussions. But there are also a few remarks I want to make on Hilbert's alleged instrumentalism.

Part of the problem, undoubtedly, lies in this word 'instrumentalism'. According to Detlefsen:

> . . . instrumentalism with regard to a given body T of (apparent) theorems and proofs [consists] in the belief that the epistemic potency of T (i.e. the usefulness of items of T as devices for obtaining valuable epistemic attitudes toward genuine propositions of some sort) can be accounted for without treating the elements of T literally (i.e. as *genuine* propositions and proofs), but rather as 'inference-tickets' of some sort.[93]

But what is it to treat the items of T 'literally'? If this means, treat them as *meaningful*, then we need to know what notion of meaning is involved. Detlefsen does not tell us. But now, suppose someone asks me if I believe - have a 'valuable epistemic attitude to' - the proposition that, if k is a regular and uncountable cardinal, and if F is a normal filter over k that contains all final segments $\{a: a_0 < a < k\}$, then F contains all closed unbounded sets. If the question is asked in normal circumstances, I will say, Yes: this proposition is provable in **ZF**, and I can point to a proof.[94] In a certain kind of philosophical context, though, I would feel obliged to give a more guarded answer, for on a certain (philosophically loaded) notion of belief, associated with a certain (philosophically loaded) notion of meaning, I would want to give the answer, No. For I do not think that transfinite set theory has, could have, or needs a thick interpretation. What this shows is that understanding talk of treating this or that literally, talk of 'epistemic attitudes' etc., awaits clarification of a great many philosophically controversial issues.

Now, it is a definite mistake to interpret Hilbert as holding that any part of mathematical language is meaningless unless a particular (philosophically loaded) notion of meaning is

---

[93] Detlefsen [1986] p3.
[94] Jech [1978] p60-61.

already in place. To be sure, Hilbert talks constantly of his 'ideal elements' having no meaning, but this talk is also constantly qualified with such phrases as 'in themselves'. Frege, remember, sometimes talks of singular terms as 'meaning something only in the context of a Thought'. No-one reads Frege as claiming that singular terms are meaningless. It is quite true, of course, that Hilbert was deeply opposed to what he believed to be Frege's demand for meaning in mathematics, and he certainly denied that mathematical theories were meaningless in that sense. But that does not commit him to the view that any part of mathematics is meaningless, simpliciter.

There is, of course, an important truth lurking behind the instrumentalist misinterpretation of Hilbert. In discussing the Master Argument, we saw that the procedure that eliminates ε-terms from proofs induces an interpretation of the ε-terms that occur in a given proof, but does not have the consequence of associating with any formula involving ε-terms a fixed interpretation. In general, the interpretation induced by the elimination method will vary as the formula occurs in different proofs. This tells us quite precisely the sense in which Hilbert thought his 'ideal elements' were meaningless. In much the same way, one might misleadingly claim that token-reflexives, say, are meaningless, intending merely to convey that they differ from proper names in carrying no fixed and invariant reference.

Against this misleading impression, though, notice that Hilbert does give at least an informal semantics for the ε-operator. To be sure, he gives no more than that: Hilbert is really only interested in exploring its syntactic properties. But still, he does give an informal semantic gloss, in which he explains it as a choice-function. In the context of arithmetic, the ε-operator can be taken to denote a minimization operation - that is, an operator that forms a definite description from an open sentence. It is extremely misleading to describe such an operator as meaningless - indeed, it is even a mistake to think that it lacks a Fregean sense.

It is worth reiterating the point made above that *some* of Hilbert's followers, in particular Genzen, understood him to be offering a kind of semantic account of infinitary mathematics, but one which avoided the invocation of objects as semantic values of infinitary mathematical notions.[95] This is not quite the way one ought to put the point, but it is nevertheless on the right tracks. Hilbert was trying to explain the meaningfulness of

[95] See Genzen [1938], English translation in Genzen [1969], especially pp247-251. Genzen's own account of the meanings of the logical constants is very much in this genuinely Hilbertian tradition.

transfinite mathematics, but in an internalist kind of fashion, focussing on the way that particular operators - term-forming operators in particular - were actually used. There is in Hilbert the suggestion (the *suggestion*) of a rule-based approach to mathematical meaning. To be sure, any attempt to work out the approach in even the most sketchy sort of detail would have to press philosophical questions which Hilbert never answers, since they concern matters very remote from his professional concerns. Nevertheless, the sympathetic interpreter ought to notice that Hilbert has copious room for maneuver in this area.

There is no reason to assert, therefore, and every reason to deny that Hilbert would have been ill-disposed to the interpretability of mathematical theories in the thin sense. There is no reason to foist instrumentalism on him, for there are many, many ways of understanding semantic notions as applied to mathematical language that lead neither to instrumentalism nor to the 'Fregean' theory that lies at the other extreme.

Let us set exegetical issues aside. Now, the most commonly identified weakness of instrumentalism, whether in the philosophy of mathematics or the philosophy of science generally, is the apparent inability of the instrumentalist to account for the effectiveness of the instrument. Dag Prawitz puts the point as follows:

> A reasonable foundation of mathematics cannot treat the transfinite part of mathematics as an instrument, a black box, that happens to give correct results; the weakness of such an instrumentalistic position . . . is obvious since the foundational task must be to explain why the instrument works, i.e. to understand it . . . In short, to make Hilbert's program at all credible, one must require that it yields an interpretation of also the ideal sentences.[96]

Now of course, this is not at all plausible if 'interpretation' is taken in the thick sense. Whatever is meant by explaining how a mathematical theory 'works', you can certainly do it without supplying that theory with an interpretation in the thick sense. There is such a subject as model theory. So this objection does not arise for the Hilbertian position I have been sketching.

Detlefsen, however, seems to hold that all interpretation must be interpretation in the thick sense. In attempting to rebut the Quinean indispensability argument, for example, he complains that the argument . . .

---

[96] Prawitz [1981] p268, and Kreisel [1964].

> . . . treats the mathematical part of science as being on an epistemic and semantical par with the non-mathematical part. Both mathematical theorems and physical hypotheses are treated as genuine propositions whose epistemic role is taken to be determined by their evidentness as truths. But for what reason? Surely applied mathematics would be just as reliable a guide to empirical truth if it were merely empirically sound or empirically conservative as it would be if it were literally true. So, in order to account for its utility in constructing successful theories, we need not ascribe truth but only truthfulness to mathematics. And we can do this without assigning to it any literal semantical status.[97]

As criticism of the Quinean indispensability argument, this is worthless, since it draws on a distinction between being 'merely empirically sound' and being literally true which is quite foreign to Quine. Much more importantly, though, the notion of 'literal truth' being appealed to here is exactly what the instrumentalist needs to be explaining, if his arguments are to be effective against anything other than a man of straw.

Nevertheless, the conviction that all interpretation of mathematical theories must involve this pernicious 'literal' meaning induces Detlefsen to attempt a reply to the Prawitz point from a genuinely instrumentalist perspective. As he understands it, the central challenge is to provide an account of the instrument's *perspicacity* (i.e. its success at delivering lots of truths about finitary mathematics - classical mathematics is the instrument, remember) and its *reliability* (i.e. its capacity to deliver *only* truths of finitary mathematics). Now, there is undoubtedly some point of contact with Hilbert's own thought here, since Detlefsen has just described, in his own terms, the problem to which the Conservation Program is intended to provide the solution. And this means that, for Detlefsen, the incompleteness theorems pose an enormous challenge, since they both purport to provide finitary truths that cannot be finitarily proven - that is, they purport to show that the ideal instrument must be either unreliable or defective in perspicacity. Having set things up in this way, then, Detlefsen must contest the common understanding of what Gödel showed. And as we shall see in **Chapter Three**, he does.

---

[97] Detlefsen [op cit] p25.

CHAPTER TWO:
Finitism, Mathematical Objects, and Intuition

Introduction: In this chapter, I attempt to discharge a number of debts incurred in **Chapter One**. In particular, I try to explain what finitary mathematics is, why one should take a realist attitude towards it, and why it is of particular philosophical interest. The chapter falls into three parts. In the first two, I characterize, but then reject, two respects in which finitary mathematics might be thought (and has been thought) to have some particular philosophical interest. In the third part, I introduce and defend my own alternative characterization of the special status of finitary mathematics.

One cannot overemphasize, in my view, the importance of the finitary/ideal distinction. It is the single most striking feature of Hilbert's Programme, and the philosophical position associated with it. If we do not understand this distinction, we cannot understand Hilbert's Programme. And if we cannot provide some convincing philosophical rationale for the partitioning of classical mathematics into finitary and ideal parts, the apparent philosophical interest of Hilbert's Programme will turn out to be illusory. In order to understand the distinction, a mere list or description of the finitary part of classical mathematics cannot suffice. One cannot merely to *stipulate*, for example, that PRA is finitary, and leave it at that. For what we need to know is *why* PRA is finitary, and why Euclidean geometry (say) is not. One might say, finitary mathematics is that part of classical mathematics which does not 'involve' infinitary objects, or infinite quantities, or continuity. But this will hardly do as a characterization. For example, it is quite unclear what counts as 'involving' infinitary objects - indeed, it is quite unclear what 'infinitary objects' are supposed to be. **PA**, if modelled in set theory in von Neumann's way, 'involves' only hereditarily finite sets; yet the axioms of **PA** are satisfiable only in domains which are at least denumerably infinite. Does **PA** therefore 'involve' infinitary objects? Does **PA** belong to finitary, or ideal mathematics? So far, we have little to enable us to get a grip on these questions.

Nor will it do to say, simply, that all and only those parts of classical mathematics that everyone accepts, including the intuitionists, should count as finitary. For Hilbert's purposes, matters cannot be left at that. To begin with, most constructivists, including all intuitionists, accepted parts of classical mathematics that Hilbert certainly counted as ideal; so if our concern is with what *Hilbert* meant by 'finitary mathematics', this answer is

extensionally incorrect. More importantly, though, mere *de facto* acceptance by all (late nineteenth century?) mathematicians seems to me too local, and too fragile, a criterion for Hilbert's foundational purposes. It is too local, in that the parts of mathematics that enjoy *de facto* acceptance by all mathematicians have in fact varied widely, even in recent history. And it is too fragile in that it appears merely to await passively the next innovations in the more foundational parts of mathematics - the kind of innovation of which Cantorian set theory is the paradigm. The Hilbertian insistence on formalization of a mature theory, followed by a proof of consistency, was intended to provide mathematics with a general means of assessing the *acceptability* of creative innovation of exactly this kind. The *finality* to which Hilbert's Programme aspires, whatever that comes to exactly, requires that a finitary consistency proof have a normative force beyond that provided by the mere *de facto* acceptance of finitary reasoning.

Moreover, if *no* principled distinction between finitary and ideal mathematics can be drawn, then Hilbert's Programme, at least in the form found in Hilbert's own work, simply collapses - quite independently of any technical objections stemming from the incompleteness results. For the Programme is completely dependent upon an alleged *asymmetry* between ideal and finitary mathematics - an asymmetry in virtue of which the former may be taken to license the use of the latter, but not vice versa.

However, it might be thought that the required asymmetry could be established without any entanglement in philosophical controversy. After all, it is a familiar point that 'constructive' proofs often provide *more information* than non-constructive proofs. Since finitary proofs are certainly constructive, it might be thought that a sufficient motivation for seeking a finitary proof of consistency is provided by the point that such a proof is likely to provide more information - about the complexity of proofs within the theory for which consistency is proved, for example - than an ideal proof of consistency can.

There is no denying that this does indeed provide a motivation for seeking constructive consistency proofs for mathematical theories, even when non-constructive proofs have been found. Indeed, some believe this to be the only genuinely compelling motivation a project such as Hilbert's can have.[98] And Hilbert certainly attached very great importance

---

[98] In contemporary usage, proof theoretic research motivated solely by considerations of informativeness with respect to, say, length and complexity of proofs is known as non-reductive proof theory. What remains of Hilbert's Programme is then classified as reductive proof theory. See e.g. Prawitz [1981] p235-236.

to informativeness considerations in motivating the search for constructive proofs. But Hilbert's project, as Hilbert himself understood it, cannot be motivated in this way. For to begin with, constructive proof is a far more capacious notion than finitary proof: any textbook of classical analysis will contain many constructive proofs which are certainly not finitary in character. Furthermore, it is by no means the case that *finitary* proofs typically yield more information than ideal proofs of the same results. One need look no further than the notorious computer-assisted proof of the Four Color Theorem for a pertinent illustration of this point.[99] The lingering dissatisfaction with this proof felt within the mathematical community has little or nothing to do with the issue of whether or not the use of a computer to provide an unsurveyable proof makes the theorem empirical in character, much philosophical agitation to the contrary. Rather, the dissatisfaction has its origins in the fact that the theorem, when established in this way, yields much *less* information than one would expect to be able to gather from an ideal, but humanly surveyable proof. Unable to follow the reasoning, we are unable to see clearly where, and how, the proof fits in with everything else that we know. And of course, this feature of (some) constructive proofs provides part of the motivation for Hilbert's *opposition* to Kronecker's insistence that only finitary (or constructive) proofs can be of mathematical value. Finally, it is clear that the *kind* of importance Hilbert attached to finitary proofs of consistency requires an altogether stronger motivation than that provided by the desire for maximally informative proofs. Someone who is moved by the desire for informativeness need not think anything fundamentally amiss with a theory for which no finitary consistency proof can be given. Hilbert, on the other hand, thought that the absence of such a proof cast doubt on the mathematical status of a theory.

A final point touching the importance of Hilbert's justificatory remarks on finitary mathematics, the most important of all for our purposes, is this. Unless we are clear about what the special characteristics of finitary mathematics might reasonably be supposed to be, we cannot be clear about what would count as a successful completion of Hilbert's Programme. Consider for example Genzen's proof of the consistency of arithmetic. There has been much debate over whether this should count as a partial realization of Hilbert's Programme. On one side, the 'vizualisable' character of induction up to ε_0 is counted in favour of a positive answer.[100] On the other side, the fact that this induction is in some (rather obscure) sense 'stronger' than the means of proof available in arithmetic itself is

---

[99] See Appel, K.I, and Haken, W. [1976]. For more details, and more discussion, see Saaty, T.L, and Kainen, P.C. [1977].

[100] See, for example, Takeuti [1987] pp86-101, and compare Pohlers [1990] pp75-76.

thought to suggest a negative answer. How one reacts to this debate will plainly be determined by one's beliefs about what, if anything, is distinctive about finitary mathematics.

What is needed here, then, is something distinctively philosophical in character. We need not only a criterion which will enable us to identify perhaps novel mathematical arguments as finitarily acceptable or otherwise, but also a compelling defense of the claim that the part of mathematics isolated by the criterion *ought to be* acceptable to any mathematician, including the many possible varieties of constructivist. If, for example, the claim is that finitary mathematics is particularly obvious, particularly well-founded, or in some other way *epistemologically* special, then we shall need to be persuaded that finitary mathematics, and only finitary mathematics, really does have this status. As we shall see, it is very difficult to defend any such claim.

Finally, let me emphasize that it is by no means obvious where the finitary/ideal boundary is to be drawn, at least if one wishes to keep faith with Hilbert's own position. Hilbert never clarified the required distinctions to his own satisfaction (or anyone else's), and there is some evidence that the discovery of the incompleteness theorems tempted him, towards the end of his life, to allow as finitary systems of mathematics much stronger than **PRA.**[101] My own view is that taking this option destroys most of the philosophical interest of Hilbert's Programme, so it will be important for me to give some justification for favoring more stringent standards.

Section One: Three Alternative Accounts of the Special Status of Finitary Mathematics. In passages scattered throughout Hilbert's writings, we find remarks about the 'concrete' and 'intuitive' character of finitary mathematics, about its 'reliability', 'clarity', and - most revealingly - its 'indispensability for all scientific thought'. These remarks, quite plainly, are intended as contributing towards the needed philosophical justification of the special status accorded to finitary mathematics in Hilbert's Programme.

---

[101] According to Bernays, it was not until Gödel's 1933 modelling of classical arithmetic in intuitionistic arithmetic that the Hilbert school realized there was a distinction to be drawn between intuitionistic methods and finitary methods as they had hitherto understood them (see Bernays [1934] p271 e.g.). Since classical arithmetic *cannot* be modelled in **PRA**, it became evident that Hilbertian finitism was more restrictive than intuitionism, and this provided some motivation for adopting a more relaxed understanding of the finitary/infinitary distinction. In Bernays' own development of the Hilbert Programme (continued in the work of Feferman e.g. - see Feferman [1964]), free use is allowed of methods which are, in my view, quite plainly not finitary in character. Cf. the remarks by Simpson in Simpson [1988], pp352-353.

But these remarks are very vague. Under different possible interpretations, they suggest that finitary mathematics has special status for the following three quite different reasons.

(A) The use of the word 'concrete', along with the insistence that finitary mathematics deals only with 'expressions', 'the concrete signs themselves', might seem to imply that finitary mathematics is supposed to be *ontologically* special. The suggestion then appears to be that the *objects with which finitary mathematics deals* are more readily accessible to us, perhaps more akin to the objects dealt with by the physical sciences, and less 'abstract' than the objects of the rest of classical mathematics.[102]

(B) On the other hand, the remarks about the 'obvious' and 'intuitive' character of *finitary mathematical truths* suggest that the special status of finitary mathematics is *epistemological* in character. Finitary mathematical truths would then count as more secure, more certain, better grounded, or something of that kind, than truths involving the infinitary parts of mathematics. This interpretation has been very influential in philosophical discussions of Hilbert and Hilbert's Programme.[103]

(C) Finally, there are many suggestions in Hilbert's writings of a vague, somewhat Fregean-sounding doctrine according to which finitary mathematics subsumes that part of mathematics without which 'scientific' thought would be impossible. (Of course for Frege, almost *all* of classical mathematics has this character in virtue of the identity of 'arithmetic' and logic.) This suggestion appears quite clearly, for instance, when Hilbert writes

> . . . as a condition for the use of logical inferences and the performance of logical operations, something must already be given to our faculty of representation, certain extralogical concrete objects that are intuitively present as immediate experience prior to all thought. If logical inference is to be reliable, it must be possible to survey these objects completely in all their parts, and the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects, as something that neither can be reduced to anything else nor requires reduction. This is the basic philosophical position that I consider requisite for

---

[102] It is notoriously difficult to give a satisfactory account of the abstract/concrete distinction. Indeed, it is difficult to give a convincing argument for the existence of some such distinction: Hilbert, as we have have in effect already seen, was sceptical about such arguments. I shall take it that the relevant feature of the 'abstract' objects of classical mathematics, platonistically understood, is their exclusion from the causal nexus. The standard epistemological puzzles surrounding platonism, after all, stem from the difficulty of understanding how objects which are not to be met with in space and time can nevertheless be cognized by creatures whose cognitive capacities, mathematical capacities excepted, appear to be best understood in information-processing terms.

[103] For more on this topic, see **Appendix One**.

> mathematics and, in general, for all scientific thinking, understanding, and communication.[104]

And again:

> The formula game that Brouwer so deprecates has, besides its mathematical value, an important general philosophical significance. For this formula game is carried out according to certain 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 our understanding . . ..[105]

Of course, it is not being implied that these three characterizations of finitary mathematics are incompatible, or even in any way in tension with each other. They need not be. However, they appear to be independent of one another, in the sense that finitary mathematics might very well satisfy any one of them without satisfying the others.[106]

I shall consider these three possibilities in order, rejecting (A) and (B) (in **Section Two** and **Section Three**), and endorsing a version of (C) (in **Section Four**). I shall also argue that nothing Hilbert actually says *forces* the ontological or epistemological interpretations on us. I think that there is reason to believe that (A) and (B) reflect aspects of his understanding of finitary mathematics that are at best peripheral.

Section Two: Finitism and Mathematical Objects: Hilbert's explanations of the infinitary/finitary distinction contain opinions concerning the nature of the *objects* considered in finitary mathematics. To a very crude first approximation, the view Hilbert appears to be advancing is this

(Ont) Finitary mathematics is ontologically committed to *expressions* only - what Hilbert calls 'the concrete signs themselves'.[107]

The intended contrast, of course, is with the 'infinitistic' objects associated with ideal, infinitary mathematics, or indeed with ordinary arithmetic, as Platonistically construed. In my view, however, finitary and ideal mathematics cannot be contrasted in this ontological

---

[104] Hilbert [1925] p376.
[105] Hilbert [1927] p475 - emphasis in the original.
[106] I say 'appear to be' simply because they are, at they stand, too vaguely expressed to permit a more confident judgment.
[107] See, for example, Hilbert [1925] pp376-377, or Hilbert [1927] p465, p469-70.

way: nor does Hilbert really think that they can. To begin to see why, we must look at **(Ont)** more closely.

(Ont) claims that finitary mathematics *does* have distinctive ontological commitments - but only to expressions, the 'concrete signs themselves'. However, some further thought will show that this is misleading. As Hilbert understands it, this commitment to 'expressions' cannot be a case of commitment to any distinctive kind of entity at all. To see why, we must review some things Hilbert says about the 'subject matter' of finitary mathematics.

First, consider the following passage from Hilbert's most famous article:

> Kant already taught - and indeed it is part and parcel of his doctrine - that mathematics has at its disposal a content secured independently of all logic and hence can never be provided with a foundation by means of logic alone; that is why the efforts of Frege and Dedekind were bound to fail. Rather, as a condition for the use of logical inferences and the performance of logical operations, something must already be given to our faculty of representation [in der Vorstellung], certain *extralogical concrete objects* that are intuitively [anschaulich] present as immediate experience prior to all thought. If logical inference is to be reliable, *it must be possible to survey these objects completely in all their parts, and the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects, as something that neither can be reduced to anything else nor requires reduction.* This is the basic philosophical position that I consider requisite for mathematics and, in general, for all scientific thinking, understanding, and communication. And in mathematics, in particular, *what we consider is the concrete signs themselves, whose shape, according to the conception we have adopted, is immediately clear and recognizable.*[108]

Although this appears to speak of all of mathematics, the context makes it clear that Hilbert is in fact speaking of finitary mathematics - or, equivalently, of metamathematics. Here are some of the theses he asserts about this discipline:

(1) 'contentual logical inference' is guaranteed to be reliable in this domain, because

(2) the objects with which we deal in finitary mathematics are (a) concrete, and (b) surveyable: moreover,

(3) those objects are none other than the *expressions* in which mathematical theories are formulated.

---

[108] Hilbert [1925] p376, my emphasis.

(1) is the most important claim here. The reason why Hilbert asserts (1) is, however, rather more complicated than one might think. To see what it is, we need to amplify (2) and (3) a little.

Now, it is very natural to interpret these principles in the following way. According to (2) and (3), we can (in principle, as one says) actually write down an array of marks - physical objects - corresponding (in a sense which can be made precise) to any given atomic sentence of finitary mathematics. The array corresponding to the standard decimal representation of the number three, for example, might be '/ / /'. Now, this is essentially the claim finally made precise, and vindicated, by Gödel's arithmetization of syntax, together with Quine's demonstration that arithmetic can be modelled in syntax.[109] Given this interreducibility, the theory of syntax and elementary number theory are mathematically equivalent. Thus, the theory of syntax may treat the expressions of any theory as *numerals* without any loss of generality. The array '/ / /' is simply the numeral for the number three in a monadic notation. The result of writing down such an array will be a complex token inscription - in the case of an atomic sentences of number theory such as '1<2', for example, the inscription might consist of the two arrays of strokes '/' and '/ /', separated by a gap. Any such token inscription can be 'surveyed', again in a sense which can be made mathematically precise. The original atomic sentence of finitary mathematics will be true just in case the second array is 'longer than' the first, where, once more, 'length' has a precisely specifiable mathematical sense. And as Quine shows, the entire content of finitary mathematics can be modelled in this way. The procedures for constructing these models of finitary mathematical assertions are (as we would now say) effective: they can be fully specified, they require no exercise of 'intuition' of any kind, and they can be guaranteed to produce a result after only finitely many steps.

Now, (1) turns out to be slightly misleading in one respect, which is this. Logical (and mathematical, of course) operations on these arrays must be guaranteed to produce only surveyable arrays from surveyable arrays. That is to say, the entire theory must be guaranteed to have an hereditarily finitary character. Some logical operations violate this constraint. Modulo this complication, though, the line of thought we are pursuing sees Hilbert as asserting (1) just because he thinks he can show, via (2) and (3), that there can

---

[109] The arithmetization of syntax first appears in Gödel [1931]. For the modelling of arithmetic in syntax, see Quine [1946].

no more be contradictions in finitary mathematics than there can be contradictions in physical reality. For the theorems of finitary mathematics can (in principle) be demonstrated in *physical models* - they can be modelled as arrays of 'expressions'.[110]

This interpretation of Hilbert is not without some initial plausibility. To evaluate it, both as interpretation and as doctrine in its own right, we must now try to see, in more detail, just what these claims about physical models are supposed to mean.

In one of the few passages in which Hilbert goes into more detail about the elements of finitary mathematics, he writes

> In number theory we have the numerals
>
> 1, 11, 111, 11111,
>
> each numeral being perceptually recognizable by the fact that in it 1 is always again followed by 1 (if it is followed by anything). These numerals, which are the object of our consideration, have no meaning at all in themselves. In elementary number theory, however, we already require, besides these signs, others that mean something and serve to convey information, for example, the sign 2 as an abbreviation for the numeral 11, or the numeral 3 as an abbreviation for the numeral 111; further we use the signs +, =, >, and others, which serve to communicate assertions. So 2 + 3 = 3 + 2 serves to communicate the fact that 2 + 3 and 3 + 2, when the abbreviations used are taken into account, are the same numeral, namely, the numeral 11111. Likewise, then, 3 > 2 serves to communicate the fact that the sign 3 (that is, 111) extends beyond the sign 2 (that is, 11), or that the latter sign is a proper segment of the former.[111]

Now, I think it is quite clear that the above passages, along with the similar passages that occur elsewhere in Hilbert's writings, do not force us to the view that we must understand these arrays of strokes as the *references* of the primitive terms of finitary mathematics. On the face of it, it is perfectly compatible with all that Hilbert says that the relationship between the arrays of strokes and the symbols of finitary mathematics should be one of *instantiation*, rather than reference. It still might be true, for example, that the 'objects of our consideration' in finitary mathematics are arrays of strokes, if the content of finitary mathematics is in some way instantiated in those arrays.

Now, whilst I think I know what Hilbert is trying to get at in this passage, I do not think that the passage makes much sense as it stands. I do not understand, for example, what it

---

[110] With some trepidation, I am inclined to see Charles Parsons as interpreting Hilbert in this way. See the references in footnotes 115 and 116 below.

[111] Hilbert [1925] p377.

means to say that these numerals 'have no meaning at all in themselves'. Presumably, numerals are signs. What would it be for a sign to have meaning in itself? No-one could seriously suggest that numerals are *natural* signs for numbers, in the sense that, say, certain cloud formations are natural signs of rain. Nor do I understand how a mere abbreviation for a sign that has no meaning in itself can have meaning. I am not sure what is being used and what mentioned, nor am I sure whether types are being talked about, or tokens. I am puzzled as to why the ideogram '1' has been selected for this starring role in arithmetic, and I am all the more puzzled since relegating this ideogram to its proper place seems apt to refute much of what Hilbert asserts towards the end of the passage - for example, the Roman numeral for the number four does not 'extend beyond' the Roman numeral for the number three.

For all the obscurity in this passage, though, I do think that I have a reasonable idea as to what Hilbert means, and I think that what he means is interesting enough to warrant spending some time trying to get at it. When we have done so, I think it will be apparent that the claim that finitary mathematics may be thought of as that part of classical mathematics that has models in arrays of strokes has very little plausibility. The way to proceed, though, is to abandon Hilbert exegesis for the moment, and review some very simple facts about counting.

Hilbert claims, then, that the 'objects of consideration' in finitary mathematics are *expressions* - the 'concrete signs themselves'. In number theory, for example, those expressions would be *numerals* (along with expressions denoting operations on numerals etc.) What are numerals? This is a bad question, and we must see why.

Consider the process of *making* and *keeping a tally* . For example, a shepherd might make a tally of her flock of sheep, by pairing off the sheep with some pebbles which are then kept safely stored in a bucket. The standard choice of verbs here is interesting. Tallies are *kept* - that is one reason why we say the shepherd is keeping a tally of the sheep using the pebbles, not keeping a tally of the pebbles using the sheep. Tallies are *made*, or perhaps we should say, *constructed.* That is another reason. Sometimes 'constructed' has its literal sense, but even when nothing is actually moved around or physically manipulated in any way, tallies have to be 'built up' out of 'constituents' - constituents which are themselves tallies, notice. The ability to keep, or make, a tally just is the ability to perform some such construction procedure.

A tally, then, is an aggregate, an array of objects, standing - as we would put it - in one-one correspondence with the aggregate of objects of which it is a tally. In any tallying procedure, what counts as a tally of what - the sheep of the pebbles, or the pebbles of the sheep - is determined by purely pragmatic criteria. A little more precisely, it is determined by the *point* of keeping the tally: the tally will typically be more accessible, more stable, more portable, more permanent, that the aggregate of which it is a tally. Given the way the world is, there is often something to be said for making your tallies out of physical objects.

Two desirable properties of tallies is that they be both *portable* and *surveyable*. For a large flock of sheep, a bucket of pebbles is not an ideal tally. It might be easier to move the herd of sheep around than move the tally around, and it might be no more obvious, at a glance, how many pebbles are in the bucket than how many sheep are in the herd. A better idea, and *one which appears to be as ancient as the idea of keeping a tally at all*, would be to make use of two tallies, as follows. First, cut some notches on a conveniently short stick. Then take up an uncut stick. As the sheep exit in the morning, run your finger over the cut stick from the leftmost notch towards the right, perhaps saying 'sheep and sheep and sheep . . ' as you correlate sheep with notches. As you reach the rightmost notch, cut a notch on the fresh stick, and then go back to the leftmost notch and repeat. In this way, twenty notches, for example, can be made to serve as a tally for one hundred and nine sheep. I shall describe this kind of tally as an *abacus* tally: think of an abacus tally as a tally kept (and constructed) in a positional notation. Any 'device' for constructing an abacus tally (which may be no more than a positional notation) will count as an abacus.

There are a number of very important points to be noted about this procedure. Notice, for example, that it appears to be describable without any use of number words, or any other kind of obviously mathematical expression such as 'first' or 'last'. All that seems to be involved in keeping a tally is the ability to employ concepts of identity and distinctness amongst particulars, and that seems set fair to count as logical in nature. A second very important feature is that the abacus procedure *appears* to be mathematically much richer and more interesting than the simple tallying procedure. Ability to construct an abacus tally undoubtably demands some increase in intellectual sophistication, for there are now (at least) *two* iterative procedures involved, with different significances. For example, some experimentation with different sized tally-sticks, on which (as we would put it) different numbers of notches can be cut, can be useful in providing an intuitive grasp of some non-trivial mathematics, including some understanding of important properties of polynomials

and congruences. And of course, an abacus tally is liable to be much more portable, and more readily surveyable, than any simple tally.

But the most important fact about the simple abacus is that the *appearance* of greater mathematical sophistication masks a crucial fact. As we would now put it, any arithmetic that can be done in a positional notation can also be done in monadic notation - 'in principle', as one says.[112] A more precise version of this claim is this: *any function (from positive integers to positive integers) that is computable at all is computable in monadic notation, by a Turing machine that 'reads' (i.e. is sensitive to) only the blank and the symbol /.* And it must be an appreciation of this fact, in some sense, that serves as the *justification* of using positional notation. From the shepherd's point of view, the assurance that the new, more complex procedure of measuring the size of the herd will not lead to any mismeasurement is provided by the fact that there is an 'effective' procedure for recovering a simple tally from any (depiction of an) abacus tally, provided the basis of the positional notation is known. In this sense, the shepherd can be assured that the abacus will not lead her astray, for she can reconstruct from it a simple tally *by procedures no more complex that those used in constructing a simple tally in the first place.*

It is perhaps worth pausing to illustrate this point. Suppose, for familiarities sake, that we are dealing with a typical nursery-school abacus, with strings ordered from right to left (so that the unit measures go on the rightmost string). To recover the appropriate simple tally from this abacus, proceed as follows. (1) Empty the beads on the rightmost string, if there are any, into a bucket. (2) Find the rightmost non-empty string. (3) Remove a bead from that string, and put it in the bucket. (4) Fill every string to the right of the string from which you last removed a bead with beads. Then (5) go back to (1), and repeat the whole process until the abacus is empty. The contents of the bucket now constitute the simple tally corresponding to the tally on the original abacus.

This procedure works, of course, because it simply reverses the procedure on which the abacus was constructed.[113] Notice the following two points. First, nothing essential to

---

[112] For this terminology, and for more on this claim, see e.g. Boolos and Jeffrey [1990] pp52 ff.

[113] Suppose we come across a decimal abacus that is so big that, if we position ourselves at the rightmost string, the remaining strings vanish over the horizon. How are we to tell the difference between this abacus encoding the number zero, say, or some vast power of ten? How are we to be sure that a journey towards the horizon will never reveal a bead? The problem is not to be solved simply by insisting on a finite number of strings for any abacus, notice. For any abacus with a finite number of strings, there is a finite bound to the tallies recordable by that abacus. Thus for any such abacus A, there will be numbers n and m such that A can record n and m, but not n+m. If you want to model elementary arithmetic in tallies,

the process requires the strings of the abacus to be the same length. It does not matter whether an abacus has strings taking n beads each, or strings each taking a different number of beads, or some combination of the two. Providing the order in which the strings were filled in the construction of the abacus tally is known, the above procedure recovers a corresponding simple tally. There is therefore no need, in describing the above procedure, to say anything like 'add *nine beads* to the next string': you only need to *fill* it. To follow the procedure, you need to be able to identify a single bead as such, an empty string as such, to tell right from left, to find the nearest string to the right of a given string etc. But none of these abilities are distinctively mathematical in character. They can be exercised piecemeal by creatures who, intuitively speaking, have no mathematical concepts at all.

Secondly, the notation described here has, in effect, a zero - an empty string, perhaps surrounded by non-empty strings. But notice that nothing in particular needs to be said about the 'meaning' of an empty string. The procedure tells you what to do with *any* string when you meet it - ignore it, empty it, take a bead off it, put some beads on it, or whatever. The procedure is the same for decomposing an abacus with no empty strings as it is for an abacus with empty strings, because (of course) the procedure for filling an abacus will, but only in *some* cases, naturally leave some strings empty. Indeed, the *notation* for an abacus need not have any element corresponding to an empty string. The idea is obvious: keep two, parallel abacuses, one recording the string on which the beads are placed, the other recording the beads on that string, in the following fashion

		oo
oo		oo
o	o	oo
oo	oo	o
oo	o	
oo		
o		

---

therefore, you need some assumption to the effect that there are abacuses of every finite size. In fact, though, the only assurance that we have that an abacus can distinguish zero from large powers of ten (or whatever the base of the abacus is) is provided by the fact that abacus configurations are constructed in a finite number of steps. Thus if there is a bead off to the left somewhere, we will in fact find it after a finite number of steps. The fact remains, however, that we cannot tell, until we actually find a bead, whether or not there is a bead to be found. This is why the least search, or minimization operator is not in general finitary.

Here, the lower series of counters identifies the strings on which the counters immediately above are to be placed. Assuming a base-ten abacus, the number recorded above is therefore three million, one hundred and six. There is no need for an element 'corresponding to' an empty string. An obvious procedure turns this complex abacus into a simple abacus, from which, via the above procedure, we can recover a simple tally. It is not to be denied that there were, as a matter of historical fact, great conceptual difficulties associated with the introduction of a zero into arithmetical calculations. But equally, the way to sugar the pill is surely to explain the notation, rather than worry about the reference.[114]

An abacus gains over a simple tally in many practical respects, such as portability, and in particular surveyability. A small, manageable aggregate with a complex structure is made to encode as much information as a large, unmanageable aggregate with a simple structure. This gain is reduced somewhat, though, if it is necessary to construct a whole new abacus every time you want to make a tally of a new aggregate of objects without losing the kept tally of some other aggregate. The natural response to the difficulty is to use the abacus to *make* the tally in the first place, then *keep* a record of the final abacus configuration, thus freeing up the abacus for the next tally-making. This can be done by some variant of the notation mentioned above - a symbolism recording the abacus strings in order, and the beads on each string in the final configuration. A *drawing* of the abacus will serve. Better still - because more flexible still - is the device we now use: a positional notation associated

---

[114] The demonstration that any numerical function that can be computed at all can be computed with these very limited resources is in fact a little paradigm of Hilbert's favoured approach to the explanation of the significance of ideal mathematics. The *utility* of the abacus type of tally is obvious, for it makes readily accessible to us a range of mathematical data vastly beyond anything comprehensible in the 'idiom' of simple tallies. But the warrant for its use is nevertheless the fact that we can show, via a general argument concerning the workings of the abacus notation (equivalently, the construction of abacuses), that nothing can be done by abacuses that cannot be done 'in principle' by simple tallies. What is more, the assurance that this is so is itself something the grasping of which involves *only the kinds of ability used in constructing a simple tally in the first place.* It is the conceptual resources used in the construction of simple tallies that are required for decomposing a given abacus tally into the corresponding simple tally.)

Notice also that the enormous advantages of abacus procedures in terms of computational flexibility are naturally accompanied by 'philosophical' questions, for example, concerning the 'ideal element' (if you will) which might be thought to 'correspond' to the empty string. At the level of this kind of counting procedure at least, this is a problem which is to be addressed by a (meta)mathematical demonstration that the presence of such an element *in the notation* adds nothing essential to the mathematical content of 'theories' using the notation. There is no zero element in the 'notation' of a simple tally, whereas there often is in the notation for an abacus. Nevertheless, nothing can be done with an abacus that cannot be done with a simple tally.

with an alphabet of *words*. Equipped with this device, you can make, and keep, the tally 'in your head' as well as 'on paper'. Any polynomial

$$a_1x^0 + a_2x^1 + \ldots + a_nx^{n-1}$$

(with a, n and x natural numbers) may be regarded as a record of an abacus configuration. (Remember that, as we noted above, it is only a convenience that x, the base of the notation, take the same value throughout. This eases the memorability and cursiveness of the notation, but that is all. Provided the notation records the 'length of wire', i.e. the number base, associated with each wire on the abacus, the simple tally can be recovered automatically from the configuration recorded.)

An abacus, then, is both a device for making a tally, and a device for keeping a tally. This dual aspect of abacuses infects our discussion of tallying procedures with process/product ambiguities, which we should now try to remove. Let us do so by reserving the word 'abacus' for the instrument by means of which particular tallies may be constructed, and speaking of particular tallies, made by an abacus, as 'abacus configurations'. Notice that the innumerable procedures for constructing simple tallies are themselves nothing other than the most elementary kinds of abacus, and simple tallies the most elementary abacus configurations.

Now - to return to the point that sparked off this long excursus into tallying procedures - *numerals* are just abacus configurations. The familiar base ten number system is an abacus, along with the base two, base four, eight, sixty etc. systems that have actually been used by humans for tallying, and many possible variants that have not been so used. Simplest of all these abacuses is the monadic system, which is the instrument by means of which we construct simple tallies. Thus the numerals '1990', '10010010' etc. are abacus configurations, and given the abacus by means of which they were constructed (in particular, given the 'length of string' used, i.e. the number base) they can be effectively reduced to simple tallies. Abacuses of the kind used in the construction of numerals typically consist of a small initial sequence of words - usually n words for an n-base system - together with rules for generating the k+1'th word in the sequence from the k'th word (where k>n). Since the need may arise for arbitrarily long tallies, there will have to be some trade-off between factors such as the memorability and cursiveness of the notation, the simplicity and brevity of the rules generating the sequence, the surveyability of the abacus configurations produced, etc. (In everyday English, for example, the rules

generating the sequence of number words mimic the base ten notation fairly accurately, but quickly become unmanageable once unusually long tallies are required (witness, for example, the confused usage surrounding words like 'billion', 'trillion' etc.) For scientific purposes, a notation involving exponents is often preferred, but this notation involves some sacrifice in surveyability.)

In addition to process/product ambiguities, though, our discussion of abacuses and tallies has been shot through with type/token ambiguities, which must in turn be removed. It is surely natural to think that, if numerals are abacus configurations, they must be *types* of abacus configurations, and not tokens. Token abacus configurations are a fancier way of classifying those arrays of objects we have been describing as simple tallies. Now, these types of abacus configuration are of course abstract objects. So if finitary mathematics is committed to numerals (or, in general, to 'expressions') as types of abacus configurations, then the objects of finitary mathematics are not, on any natural understanding of those words, 'the concrete signs themselves'. Rather, they must be the types of which such concrete signs are tokens.

But we should not be too hasty about this. The question we have to ask here is, When are two token abacus configurations said to be of the same numerical type? Charles Parsons has suggested that the relation 'x is of the same numerical type as y' might be regarded as a relation holding between physical objects - in this case, the token abacus configurations.[115] If we can define this relation without quantifying over abstract types, as Parsons thinks we can, we might seem to have some prospects of vindicating a view according to which 'the concrete signs themselves' constitute the whole content of finitary mathematics.[116]

Initially, the prospects for such a definition seem promising. We described above a procedure for clearing an abacus - a procedure, apparently describable without the use of any specifically mathematical vocabulary, the implementation of which yields a simple tally from a given abacus configuration. Evidently, two token abacus configurations are *of the same numerical type* just in case the result of implementing that procedure (or some equivalent of it) on both results in simple tallies of equal 'length', where the required notion of 'length' will, of course, have to be defined. We can evidently define a notion of

---

[115] See Parsons [1971], reprinted in Parsons [1983] p53.

[116] There are passages in Parsons' writings which suggest that he thinks of Hilbert's finitary mathematics as that part of number theory which can be modelled in tallies, and also that this was Hilbert's view - see e.g. Parsons [1980], pp153-154. But the evidence is too meagre for me to be confident in attributing this view to him.

addition for tallies - in terms of concatenation, after the manner of Quine [1946]; and with addition in hand, all of the familiar arithmetical operations can be defined. On this basis, it certainly seems plausible to suggest that at least a goodly part of the content of elementary arithmetic can be captured within a theory that quantifies over token tallies alone.

But there might seem to be a fundamental problem about this. In order to say, with full generality, what it is for two tallies to be of the same *length*, we shall have to mention tallies. This prompts the question, 'What are tallies?' Well, as we have just seen at length, tallies are arrays, or aggregates, of objects. What are objects?

Thus baldly posed, this is a question which can only have silly answers. Evidently, we can't set about to look for a distinction between objects and other things. The point is, though, the question 'What are numerals' also seems sure to have only silly answers, and for the same reason. In our examples of tally-keeping, we have seen pebbles, marks on sticks, beads, token names, token inscriptions used for keeping a tally, and these are no doubt physical objects on any accounting. On the other hand, we have also seen, for instance, word-types used for keeping a tally; and these types are not physical objects. The question, 'Are tallies physical objects?', betrays a confusion. The word 'tally' does not pick out a *kind*, either artificial or natural. It does not mark a distinction amongst the furniture of the universe, a condition that an 'object' or collection of objects, whether abstract or concrete, might or might not satisfy. Rather, any (kind of) thing at all can serve as a tally (modulo some irrelevant practicalities). All this means is, anything at all can play a role in 'constructing' a tally, can play a role in the procedure of making, and keeping, a tally.[117]

Tallies, then, are rather like chess pieces (to make what has become a very well-worn, but none the less useful, analogy). To be sure, some chess pieces are indeed physical objects - the Staunton-type chess pieces now standing on my chess board are surely physical objects. What of the Staunton-type chess pieces that appear on my computer screen when I call up Chessmaster 2100? Presumably, they are also physical objects. Do the same physical objects appear each time I call up Chessmaster 2100? To be sure, pieces of the

---

[117] The analogue for tallies (and thus for numerals) of the question concerning numbers that so troubled Frege - is Julius Caesar a number? - will surely puzzle no-one. Can Julius Caesar serve as a tally? Sure. In the House of Commons, members vote by leaving the Chamber of the House and standing in one or other of two corridors, one for the Yeas, one for the Nays. The clerks count the votes by making a head count in the corridors. What the clerks are counting, notice, are *votes*, not members. But they count votes by counting heads - the members serve as a tally. Perhaps the Roman Senate voted in some such way.

same *type* appear each time. Are chess pieces to be identified with these types? What of the types exemplified by chess pieces other than the familiar Staunton variety? Presumably, all rooks (say) should count as belonging to the same type. But they are certainly not all of the same *physical* type. Given an arbitrary object, abstract or physical, no amount of examination of that object will determine whether or not it is rook. Rather, the rook is a role in the game of chess, and anything at all can play that role.

The needless puzzles lurking behind the question, Are chess pieces physical objects, also lie behind the question, Are tallies physical objects? It happens that many of our examples of keeping a tally involve keeping, and constructing, something which is on any accounting a physical object. Given the way the world is, and given what tally-keeping is for, that is not surprising. But nothing intrinsic to the practice of making, and keeping, a tally demands that the tally kept be a physical object or collection of physical objects. Suppose there are sets, in the sense in which the Platonist is said to think there are sets - non-spatio temporal somethings that we perceive by the pure light of the intellect. Then, in principle, sets will serve as well as anything else for keeping a tally. So would individual substances, souls, entelechies, or whatever.

Recall now that we entered into this line of thought by way of a claim that numerals, the 'concrete signs themselves', are the content of finitary mathematics, that finitary mathematical claims involve reference to, quantification over, only numerals. We have seen that numerals may be regarded as tallies - as types of abacus configurations. But tallies, I have now argued, are not a kind of thing: tallies are any (kind of) thing at all, only used in a certain way. But then the consequence of this, notice, is that finitary mathematics seems to have no *distinctive* ontological commitments. Finitary mathematics, on the line of thought we are developing, is committed only to tallies, and anything at all can be a tally. We might still allow, in a Hilbertian vein, that tallies can be usefully thought of as arrays of strokes, such as / / /. This can help make some finitary mathematical truths 'intuitive', in the sense in which Hilbert speaks of drawings of triangles etc. as making certain geometrical truths intuitive.[118] Indeed, if we think of tallies in this way, then finitary mathematics becomes a kind of *geometry*. But it is a confusion to think that these arrays are, distinctively, the *content* of finitary mathematics, that finitary mathematical terms ever *refer* to stroke arrays. Stroke arrays are perhaps elements of acceptable models of finitary mathematics, along with pretty much anything else: but finitary mathematics is committed

---

[118] For my views on what that sense is, see **Appendix One.**

to none of these models in particular, any more than an ideal theory such as Euclidean geometry is committed to the breadthless lengths, dimensionless points etc. of its 'intuitive' model. You are free to think of tallies in any way you choose, within the bounds imposed by the relevant axioms. The commitment to 'expressions' mentioned in thesis (Ont) above therefore dissolves: for anything at all can count as an 'expression' in the relevant sense. To be an expression is to play a role in a certain kind of procedure, and anything at all can play that role. Talk of ontological commitment to expressions is as empty as talk of ontological commitment to objects, and for the same reasons.

This might seem to provide grounds for the assertion that Hilbert's interest in what he calls 'expressions' as the 'objects of our consideration' in finitary mathematics is not, and cannot be, motivated by the thought that expressions are 'less abstract' than, say, sets.[119] For any such thesis must involve the confusion of thinking of expressions as a *kind*.

But this is still too quick. Too see why, we need to return to Parsons' claim, mentioned above, that the relation 'x is of the same numerical type as y' can be explicated as a relation amongst physical objects - numerals, in our reconstruction of the finitist position. The crucial question here is this: May we assume that every tally in which finitary mathematics is interested is of the same length as some tally which *is* an array of physical objects? A positive answer to this question would at least enable the Hilbertian finitist to maintain that finitary mathematics has no *need* of abstract objects, even if it is in some ways helpful or natural to think of tallies as, say, numeral types (rather than tokens), or indeed as sets. On the other hand, a negative answer may force a commitment to abstract objects. If we are able to regard tallies as physical objects without loss of generality, then we can commence our definition of the binary predicate 'x is of the same numerical type as y' confident that the variables can be understood to take as values physical objects (or arrays thereof), and that nothing mathematically important will be lost by this limitation.

Now, some care needs to be taken throughout the discussion of this question if we are to avoid encumbering Hilbert with onerous and unnecessary philosophical obligations. In particular, I want to emphasize that nothing in Hilbert's writings, so far as I can see, forces us to interpret him as a nominalist. To be sure, he makes remarks, such as those above,

---

[119] Here I differ from Linda Wetzel, who in a recent piece (Wetzel [1989]) groups Hilbert together with Field and Hodes as 'nominalists', for whom expressions are preferable to, say, sets on ontological grounds. Hilbert does not belong in this company, and his position on the content of finitary mathematics is untouched by her arguments.

which suggest that computations, or proofs, are to be thought of as token arrays of expressions. However, he also makes remarks that suggest that he thought of such arrays as abstract types. Nowhere does he commit himself unequivocally on the *general* issue of the existence of abstract objects, or even on the issue of the acceptability of abstract objects in mathematics (or indeed elsewhere). We have to recognize once more that, in seeking to determine the ontological commitments of finitary mathematics, we are pressing philosophical questions which Hilbert simply ignores. We are free, therefore, to adopt on his behalf any plausible answers compatible with the rest of what Hilbert believed.

The important point is this. Hilbert's finitism is *not* grounded on any general rejection of abstract objects. Hilbert's Programme is not an exercise in nominalist reduction: rather, it is an exercise in the clarification and justification of the use of the actual infinite, the infinite of classical mathematics. This exercise is certainly constrained, in Hilbert's thinking, by some weak kind of scientific naturalism, but it is not at all obvious that the kind of naturalism in question *forces* the rejection of abstract objects.[120] On the contrary, it seems perfectly compatible with everything that Hilbert believed that there should be, say, *properties* , or *relations* . Properties, understood as 'abstract objects', would be acceptable enough to a naturalist of Hilbert's kind, providing an acceptable account could be given of how we can detect those properties. If a property has instances in the physical world, and perhaps even if a property *could* have instances in the physical world, then such an account might very well be forthcoming. Alternatively, if mathematical objects include certain *structures* , for example, and if those structures can have instances in the physical world, then it is open to Hilbert to regard those structures as abstract mathematical objects, epistemic access to which goes in part (perhaps) via access to their instances. Of course, there are very large issues to be addressed before so much as the intelligibility of such an account of mathematical knowledge is to be established. My point here is simply that, *if* such an account could be given, then Hilbert need have no objection to it on the grounds that it countenances abstract objects.

Let us now return to our question: May we assume that every tally in which finitary mathematics is interested is of the same length as a tally which is an array of physical objects? A positive answer must cope with two large problems.

---

[120] See e.g. Hilbert [1925], and especially Hilbert [1917].

The first problem is this. It is clear that any argument to the effect that we can assure ourselves of the consistency of finitary mathematics by showing that any theorem of finitary mathematics has a tally-type model involves some considerable idealization of our *actual* constructive powers. The canonical verification of even so simple a theorem as, say, $2^{20}<2^{50}$ - a verification that proceeds by reducing these numerals to simple tallies, that is - is already something for which there is not world enough and time. Really very simple computations can already be such that the fastest physically possible computer could not complete them in the period between the big bang and the heat-death of the universe. It is quite unclear whether a finitarily acceptable justification of the cognitive idealizations involved in these claims about our ability to construct physical models for finitary theorems can be given. One might say, all that is required is that we accept that it is in principle possible to construct tallies of any *finite* length, thus respecting at least the fundamental constraint that idealization stop short at accrediting us with the powers to complete infinite tasks. The problem is the nature of the modality in that claim, for it cannot be that of *physical* possibility.[121] But the appeal to some non-physical modality at this point seems intensely problematic, since it seems certain to bring with it exactly the same demand for justification and clarification that Hilbert is attempting for the infinitistic parts of classical mathematics. A modalized finitism has abandoned the kind of clarity and simplicity of mathematical content that Hilbert's dialectical strategy requires.

This difficulty reappears in a still more intractable form in the second of our two problems. As things actually are, there is no reason to believe that finitary mathematics as Hilbert understands it can be modelled in *token* arrays of expressions, because there is no reason to believe, and some reason to doubt, that there are enough token expressions to be had. Token expressions are physical objects or aggregates thereof - inscriptions, disturbances of the air, or whatever - and are therefore bounded by the resources of the physical world. There is some reason to believe that those resources are finite. Indeed, there is some reason to believe that it is a matter of *natural law* that those resources be finite - the laws of physics themselves may dictate that the quantity of matter in the universe must lie below a finite bound. If that is so, then only some version of *strict finitist* mathematics can be thought of as having physical models. A distinctive feature of strict finitist mathematics is that the natural numbers are not thought of as closed under the elementary arithmetical operations. Thus in a strict finitist arithmetic, the existence of natural numbers n and m need not guarantee the existence of a natural number n+m. And of course, any insistence

---

[121] For a discussion of the difficulties of interpreting modal claims such as this, see Kessler, G..[1978].

on physical models of finitary mathematics forces this conclusion, if the universe is finite. The existence of resources sufficient for the construction of arrays corresponding to n and m individually does not guarantee the existence of resources sufficient for the construction of their sum.[122]

Hilbert was not a strict finitist, and there is reason to believe that he would have regarded strict finitism with distaste. We cannot, therefore, limit finitary mathematics as Hilbert understands it to that part of mathematics that can be provided with physical models. Is there any sense, therefore, in which we can still think of tallies as arrays of physical objects?

A natural thought here, I think, is the following. Even if it turns out that, for some token tally n, we lack the resources necessary to construct from n the token tally $\ulcorner n' \urcorner$ (the successor of n) - indeed, even if it is a law of nature that this is so - there must be *some* sense in which it is possible to construct the successor of n for any given tally n. Natural as this thought is, though, we have already seen that the modality here cannot be that of physical possibility. It must therefore be that of what Plantinga calls 'broadly logical' possibility.[123] The consequence of this proposal, then, would be that at least some tallies must be thought of as being of the same length as some possible, but non-actual, physical objects. There would be more tallies - in fact, denumerably many more - than there are actual physical objects.

This does not immediately preclude all those non-actual tallies counting as physical objects, notice. On at least one prominent account of broadly logical possibility, that advocated by David Lewis, all the objects that there are are physical objects, although some of those physical objects are not *actual* physical objects - that is to say, there are physical objects that do not exist in *this* world. But once again, this way of ensuring material enough for the construction of all the tallies in which arithmetic is interested seems certain to be deeply unattractive to anyone who in the least bit sympathetic to the outlook of Hilbert's finitism. No-one willing to swallow all those possible worlds will be likely to choke on real

---

[122] In fact, not even strict finitist mathematics can be thought of as having physical models in this way. Although the strict finitist will hold that the natural numbers are not closed under the elementary arithmetical operations, he will also most likely hold that we cannot, even in principle, know where those operations begin to fail. Thus although the strict finitist will be committed to the existence of natural numbers n and m such that no natural number n+m exists, he will also say that we cannot produce any examples of such numbers. This position is therefore also incompatible with the existence of a bound imposed by limitations which presumably have a physical specification.

[123] See Plantinga, A., [1974], pp 1-2.

numbers. Beside the problem of accommodating Lewis's ontology within a mildly naturalistic theory of our epistemic capacities, the problem of so accommodating classical analysis seems minor indeed.

An alternative - one which has also been canvassed in the literature in response to problems of this kind - would be to attempt the construction of tallies out of, say, space-time points.[124] For all that you can't bounce a baseball off them, space-time points count as physical objects at least in the sense that the existence of such things is required by the truth of physics (as we now have it); and, given that space-time is said to be continuous, this certainly ensures an adequate supply of material. Once again, though, this is a route which is closed once we accept the limitation to the finitistically acceptable, for it simply assumes notions, such as the 'intuitive' continuity of space and time, which are on all fours with those for which Hilbert is seeking to clarify and justify. As Hilbert [1925] makes abundantly clear, the continuum of space-time points is a paradigm of an ideal notion.

Simple as they are, these objections seem to me to be decisive against the view that tallies can be taken to be arrays of physical objects, compatibly with the basic tenets of Hilbert's project. The second objection in particular compels us to acknowledge that there are tallies which are important to finitary mathematics (countably many of them, in fact) but which are longer than any tally that is an array of physical objects. There are, no doubt, ways around this problem, but they are apparently closed to Hilbert. He cannot think of finitary mathematics as that part of mathematics that has physical models, and the special status of finitary mathematics cannot be defended on the grounds that finitary mathematics can be modelled in arrays of expressions.

It seems to me necessary, therefore, to take the tallies with which finitary mathematics deals to be abstract objects, and be done with it. This conclusion is forced, though, not by the *nature* of the objects of finitary mathematics, whatever that might mean, but rather by their *number*. We will therefore abandon the attempt to construe the relation 'x is of the same numerical type as y' as a relation holding amongst physical objects. At best, it holds between abstract objects, tally-types. We now need to know more about these abstract objects.

---

[124] The locus classicus of this approach is Goodman, N, and Quine, W.V.O. [1947].

Initially, the retreat from token arrays as tallies to type arrays as tallies might still seem to enable us to think of at least a great many tallies as *types of physical object*. Indeed, this seems like mere common sense: if two shepherds make each make a tally of the herd, using abacuses of the same type, the result will be two abacus configurations which are of the same type, and it is of course the type that we are interested in, not the tokens. If numerals are types of abacus configuration, then they are of course abstract objects. Nevertheless, at least some of them will have (actual) physical tokens. A story will have to be told about those that do not, and those that cannot have physical tokens - a story which will presumably appeal to the process under which abacus configurations are constructed. That is to say, it will be a story about *abacuses*, about abstract construction routines. If the story turns out to be a story about possible constructions, a form of the first of the objections we have been discussing will arise all over again.

But prior to that problem, there is a fundamental difficulty facing any attempt to construe the content of finitary mathematics in terms of types of tally. The token arrays

(a) //////// (b) 00000000 (c) ########

are certainly of different physical types - they differ in their manifest physical properties. Intuitively, though, they should count as tallies of the same numerical type - tallies of an aggregate of eight objects. However, the following question now arises: Shall we say that the arrays

(d) /// (e) 0000 (f) #####

are of the same *physical* type as the arrays (a), (b) and (c) respectively? Presumably not. We want (d) and / / /, say, to be of the same physical type, and (d) and (a) to be of different physical types - since we want (types of) expressions drawn from the same vocabulary, but of different length, to be of different *physical* type. The whole point, after all, is to find physical analogues for mathematical relations amongst numerals. The fact that (a) and (d) are constructed by the same procedure from the same vocabulary, then, does not suffice to ensure their being of the same physical type.

The problem, though, is that the only apparent basis for the relevant physical difference between (a) and (d) is that *the operation of adding a stroke has been iterated more often in*

the construction of (a) than in the construction of (d). This, or something equivalent to it, is in fact what we mean when we say that (a) and (d) differ in 'length', for in the arrays

(g) /// (h) / /

we must count (g) as *longer*, in the only mathematically relevant sense, than (h). (And notice, incidentally, that you can *see* that (g) is in this sense the longer array.) This relation in 'lengths', however, cannot be associated with any process of physical measurement.

It is perhaps not immediately clear what this argument shows, so let me try to explain a little further. The project, you will recall, was to identify the content of finitary mathematics with numerals, now considered as abstract types. The idea was that these abstract types can be used to model finitary mathematics. The argument above does not challenge this: rather, it shows that it is only trivially true. Numerals - expression-types of finite length, constructed from a finite alphabet - are enumerable, and, trivially, anything enumerable can be used to provide a model of finitary mathematics. More than this, though, the argument shows that the mathematical properties of arrays of numerals cannot be reduced to or associated with their non-mathematical properties without presuming notions which are essentially mathematical in character. The project of defining equality amongst numerals whilst quantifying over *types* of physical objects is bound to end in circularity, for some analogue of the notion of iterating an operation a certain *number* of times appears to be the sole adequate ground for the required notion of *physical* type.

The point here, I should stress, is not a point about the possibilities of acquiring mathematical concepts by abstraction from experience. Nor it the claim that you could not tell whether or not two objects were of the same numerical type by bringing to bear only a mastery of physical concepts. Those are separate issues. The point is that the attempt to introduce numerical concepts via quantification over physical types only, to capture the content of finitary mathematics whilst appealing to constructions involving types of physical object only, is bound to be circular.

This in turn shows, though, that the only apparent ways of showing that the objects of finitary mathematics are somehow more concrete, less abstract, more physicalistically acceptable than those of ideal mathematics lead nowhere. And notice that this conclusion is fully in accordance with Hilbert's conception of mathematics as we elaborated it in

Chapter One. The project of finding physicalistically acceptable surrogates for mathematical objects as Platonistically conceived can only be motivated by a conception of the nature of reference, of what is required for creatures like us to be able to refer to objects, that implies that reference to physical objects is somehow less problematic than reference to mathematical objects. But that is not a Hilbertian view. As we have seen, Hilbert's view is that intersubjectively manageable criteria for the rational acceptability of scientific theories must be internal to those theories. Once a theory has satisfied those criteria, no further questions about the existence of the references of the terms of those theories can intelligibly be raised. There is no further perspective from which reference to mathematical objects can be seen to be more problematic than reference to physical objects.

Since he has so often been misunderstood in this respect, let me belabor the point. Hilbert's position is to be sharply contrasted with that of an instrumentalist such as Hartry Field. In his 'Science Without Numbers', Field praises Hilbert's axiomatization of Euclidean geometry as one on which 'the quantifiers range over regions of physical space, but do not range over numbers.' This, he thinks, illustrates a general strategy for getting rid of apparent reference to abstract objects in mathematical theories, via a 'representation theorem' that enables one to 'find abstract counterparts of concrete statements' and vice versa. He writes

> Consequently, premises about the concrete can be 'translated into' abstract counterparts; then, by reasoning within [the abstract mathematical theory], we can prove abstract counterparts of further concrete statements, and then use the homomorphism to descend to the concrete statements of which they are abstract counterparts. The concrete conclusions so reached would always be obtainable without the ascent into the abstract . . . but the ascent into the abstract is often a tremendous saving of time and effort.[125]

But this completely misconstrues the point of Hilbert's project. Such a homomorphism does not function, for Hilbert, as a means of getting rid of reference to 'abstract objects'. It functions as a way of getting rid of any suggestion that there is a principled, scientifically manageable distinction to be drawn between 'abstract' and 'concrete' objects in the first place. On a Hilbertian view, a mathematical theory provably conservative over **PRA** has already satisfied the only defensible criteria of rational acceptability there are for mathematical theories. Whether you can interpret the objects of the theory as 'abstract', or

---

[125] Field [1980], pp24-25.

'concrete', or both, is simply irrelevant.[126] For Hilbert, the semantic notions of 'truth', 'reference' etc. are subordinate to this account of acceptability of a theory: they do not constrain it. As we shall see in more detail shortly, any systematic study of semantics, or anything else for that matter, must, according to Hilbert, presuppose finitary mathematics. The thought that such a study could then give grounds for some fictionalist reconstrual of mathematics, in virtue of having revealed that theorems of mathematical theories are not really true in the sense explicated by an acceptable semantic theory, would strike Hilbert as putting the cart before the horse. And not unreasonably, in my view.

I want now to move on to a discussion of thesis (𝔹) on page 79 above - the claim, that is, that finitary mathematics has a special status in virtue of its epistemological properties. It will turn out that the argument given above, against the possibility of specifying the content of finitary mathematics whilst quantifying over only types of physical object, is also relevant to certain attempts to defend the claim that at least some finitary mathematical truths are especially *evident*.

Section Three: the Epistemology of Finitary Mathematics. From now until the end of this chapter, we shall be preoccupied with the controversial notion of *mathematical intuition*. In **Section Four**, I shall try to defend a certain conception of mathematical intuition, and explain the role that it plays in a Hilbertian philosophy of mathematics. For the present, however, my purposes are negative. I want to show that a *different* notion of mathematical intuition, one that has often been foisted upon Hilbert, in fact plays no significant role in Hilbert's Programme, and in particular has no special relevance to finitary mathematics.

It cannot be denied that Hilbert's writings on foundations make considerable use of a notion of 'mathematical intuition'. We need to know if this notion is important to Hilbert's understanding of his central distinction between ideal and finitary mathematics, and if so, we need to know what this notion is.

The expression 'intuition' is used in a great many different ways, both in ordinary speech and in philosophical discussion. Much of this variety we may safely ignore. Sometimes,

---

126 Field is prepared to allow that there is no question whether '2+2=4' is true in arithmetic. Still, for Field, '2+2=4' is not true. This is because there is some perspective we can occupy from which truth in arithmetic can be seen as a species of falsehood. Hilbert denies the intelligibility of this suggestion. Some further discussion relevant to this point is given in the later sections of this chapter.

for example, an intuition is just a kind of *hunch*. A mathematician might, in this sense, have an intuitive conviction that the continuum hypothesis, say, is false. This usage is philosophically uninteresting and uncontroversial. The same might be said for what is now the most common usage of 'intuition' in philosophy, in which an intuitive belief is to be contrasted with a theory-guided belief. Here, intuition is something like pre-reflective thought on some subject matter, potentially revisable upon reflection.

To make our discussion manageable at all, it will be necessary to restrict our attention to what I take to be the philosophically most controversial usage of 'intuition'. There are, I think, two primary marks of this usage. Firstly, intuition in this sense is a special faculty of the mind, purportedly closely analogous to a perceptual system, and in particular to the visual perceptual system. Secondly, a belief's being intuitive, in this sense, is taken to provide an epistemic warrant for that belief, no less authoritative than the warrant provided by (deductive or inductive) argument.

Thus, if challenged to justify my belief that there is a computer screen before me right now, I might say *'I can see it'*. Modulo certain complications which need not detain us, many epistemologies recognize claims such as this as providing a very strong warrant for the rational acceptance of the belief that there is indeed a computer screen before one. In what is supposed to be an analogous way, the philosophically most controversial usage of 'intuition' allows that 'intuition' can provide this kind of strong warrant for rational belief.

The following analogy may help to focus our attention on these two features of this usage of 'intuition'. A certain kind of traditional ethical intuitionist, if asked to defend certain very basic kinds of ethical assertion, such as (say) the assertion that torturing children is wrong, was apt to say something like, *'I can just see that it is wrong'*. Here, of course, 'see' has some figurative sense, since our theorist would hold that this reply provided a warrant for the belief even if one had never actually seen children being tortured. Intuition is this figurative kind of 'sight' - a special mental faculty closely analogous to vision which provides an epistemic warrant for some very basic, 'underived' kinds of belief.

A striking feature of this special mental faculty is that it is a faculty narrowly attuned to ethical truths. The need to postulate the existence of such a faculty was prompted by the conviction that ethical knowledge possessed two large and philosophically impressive features. First, it was thought that ethical truths were not to be met with in the physical world. The perceptual capacities adequate for the explanation of our ability to grasp the fact

that Jonny was torturing the cat were thought inadequate for the explanation of our ability to grasp the fact that, in torturing the cat, Jonny was doing something *wrong*. The former fact was perceivable by the senses: not so the latter. But the *immediacy* of the ethical judgment, the alleged fact that it was not to be thought of as derived from anything ethically more basic (such as knowledge that torturing causes pain, or whatever), was thought sufficient to suggest a strong analogy with ordinary sensory perception. Just as you could literally see the torturing, you could figuratively 'see' - that is, intuit - the wrongness. Second, in addition to the conviction that ethical facts were not to be met with in the world, the need for a special faculty of ethical intuition was thought to arise from the alleged special epistemological status of at least some ethical truths. These truths were thought to be certain, absolutely reliable, unrevisable in the course of experience. Nothing that we could meet with in the world, it was said, would suffice to show that the judgment that killing children is wrong was ill-founded. In addition to being especially attuned to ethical facts, then, at least some ethical intuitions were certain, in a way in which no empirical judgments could be certain.

Now, ethics and mathematics are the two great natural sources of modal claims. As with ethics, in mathematics - on a commonsense view at least - we have a tissue of propositions unrevisable in the face of experience, and assertions which are true, but not in virtue of anything to be met with in the world. And in philosophical writing on mathematics as with philosophical writing on ethics, one sometimes meets with just this notion of mathematical intuition as a special faculty of the mind, attuned to non-natural mathematical facts and nothing else, the deliverances of which have all sorts of exciting epistemological properties. The question we must ask, therefore, is this: Is Hilbert's Programme involved in advancing some doctrine of this general kind, with respect to finitary mathematics?

No: it is not. To begin with, as we have already seen in our discussion of Hilbert's dispute with Frege in **Chapter One**, Hilbert was generally hostile to any attempts to make mathematics dependent on intuition, in the sense of direct, extra-systematic perception of truth. The major advantage of the axiomatic method as Hilbert understood it was precisely that it shifted attention away from immediate apprehension towards considerations of the overall coherence and fruitfulness of mathematical theory. A consistent theory has theorems which are true, and *therefore* has terms which refer: that *this* was the order in which the notions of truth, reference and consistency were to be explained was, for Hilbert, the fundamental feature of an acceptable epistemology for mathematics, precisely because it freed the mathematician from any reliance on this extra-systematic insight into

mathematical truth. This is one reason why Hilbert, in general, makes no attempt to select as axioms formulae which are 'obviously true' with respect to the intended interpretation when formalizing a mathematical theory. Rather, he selects as axioms formulae which facilitate the metamathematical study of the system - which has as its principle goal a proof of consistency.

If there is to be an issue here at all, then, it can only arise over the question of whether, *in the special case of finitary mathematics*, where consistency is to be established in some more direct way than via a specification of a model, Hilbert must rely on something like the controversial notion of mathematical intuition to establish consistency - whether, for example, he thinks it important that the axioms of finitary mathematics be intuitable in the controversial sense when modelled in numerals.

Now, let us be clear about one thing at the outset. It is not true that Hilbert speaks of finitary mathematics *alone* as intuitive. Intuitability *in some sense or other* is not, for Hilbert, the particular mark of the finitary. Any doubts about this should be dispelled by a glance at Stephan Cohn-Vossen's book 'Anschauliche Geometrie', published in joint authorship with Hilbert since it is in fact an expansion of Hilbert's 1921-22 Göttingen lectures on geometry.[127] A great many of the geometrical theorems and arguments there described as intuitive by Hilbert undoubtedly belong to ideal mathematics.[128] The usage of the term 'intuition' in this book is in fact typical of Hilbert's usage throughout his career. Most often, he uses the expression 'intuitive' (anschauliche) in what should be a philosophically uncontroversial way - an intuitive theorem, in this sense, is just one that is easy to accept on first acquaintance, perhaps with the use of visual aids such as diagrams, figures etc.[129]

---

[127] For some reason, the English language edition is entitled 'Geometry and the Imagination' (see Hilbert and Cohn-Vossen [1952]). Be that as it may, the English translation of the German title is 'Intuitive Geometry'.

[128] See for example the Preface to 'Geometry and the Imagination' (Hilbert and Cohn-Vossen [1952] piii). Here, Hilbert stresses the importance of what he calls 'intuitive understanding' throughout all of geometry, and freely uses the same language of 'concrete intuition' he uses in his discussions of finitary mathematics. Writing of 'the proof of the fact that a sphere with a hole can always be bent - no matter how small the hole' [loc cit], for example - surely an ideal statement - he states that the relevant theorem 'can be treated in such a fashion that even one who does not wish to follow the details of the analytical arguments, may still gain an insight into how and why the proof works.'

[129] In the same passage ([op cit] piv) he describes his presentation of geometry as 'based on the approach through visual intuition'. 'Vision' here, notice, is just ordinary vision. There is no suggestion, either here or in any other passage known to me, of a special purpose 'vision' attuned to mathematical facts.

Thus, if there is to be an issue to discuss at all, the question must be whether the other, philosophically contentious usage of 'intuitive', *also* occurs in Hilbert, in connection with the truths of finitary mathematics. If there is a case to be made here, then the best evidence for it is undoubtedly a passage which we have already encountered. The crucial parts of that passage are those emphasized below:

> . . . as a condition for the use of logical inferences and the performance of logical operations, something must already be given to our faculty of representation, certain extralogical concrete objects that are *intuitively present as immediate experience prior to all thought.* If logical inference is to be reliable, it must be possible to survey these objects completely in all their parts, and *the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects, as something that neither can be reduced to anything else nor requires reduction.* This is the basic philosophical position that I consider requisite for mathematics and, in general, for all scientific thinking.[130]

From this passage, this much at least is clear: what it is to have an intuitive grasp of a fact, in the sense of this passage, is to have a grasp of the fact 'prior to all thought'. An intuitive fact is therefore to be contrasted with a fact that is grasped (or perhaps graspable) only after some process of reflection, or the implementation of some other sort of discursive procedure of inquiry. And this, of course, immediately opens up a morass of familiar objections. Graspable by whom? By me; by Gauss; by the Martians? We all know that the things we are able to grasp 'prior to all thought' change as we get older and wiser - or slower, as the case might be. Presumably, a mathematical assertion graspable 'prior to all thought' must be one accepted by anyone capable of understanding the assertion at all. But what are the prospects of a principled account of the distinction between rejecting a mathematical assertion because one has not understood it, and rejecting it because one thinks it is false?

But this reaction is unreasonably uncharitable to Hilbert. Whatever 'intuitive' means here, we know that, for example, the Euclidean parallels postulate does not count as intuitive for Hilbert, despite the fact that ever so many people have found, and indeed continue to find, it obvious. So it cannot be any part of Hilbert's intentions to suggest that a test for intuitiveness, in the sense explained in this passage, is to be provided by some kind of survey of what mathematicians, or some other group, are prepared to say they find obvious. Rather, I think we might see him as suggesting that there is a level of cognitive activity, perhaps discoverable only after long enquiry, which is the necessary substratum of

---

[130] Hilbert [1925] p376, my emphases.

'scientific thought' - *and* that a certain minimum of mathematics is embedded in this level. By 'scientific thought', I think we may for the moment understand Hilbert to mean something like, *objective* thought - thought directed towards the representation of quantitative features of an objective, enduring, mind-independent reality. Of course, it will come as no surprise that a minimum of mathematics is to be found in thought of this kind; but Hilbert is also suggesting something about what that minimum might be, and why it should count as basic.

In describing this level of cognitive activity as a *substratum* of 'scientific thought', I mean to be suggesting only that ascription to an agent of the capacity for any 'scientific thought' whatsoever implies ascription of mastery of the conceptual repertoire located at this level. That repertoire must include, Hilbert seems to suggest, at least the ability to deploy, synchronically and diachronically, criteria of identity and distinctness amongst particulars - the ability that is basic to the construction of a simple tally, and therefore basic to anything we could make sense of as a practice of counting. Closely related to the ability to make such distinctions is the ability to deploy certain relational concepts, such as 'longer than', 'farther from', 'the same size as', 'bigger than'. In short, the conceptual resources available at this minimal level will include something akin to the concept of a well-founded linear order.

In addition to the second of the italicized passages above, the following passage provides some additional evidence that Hilbert's view is indeed that the primitive ability to employ criteria of identity and distinctness amongst enduring mind-independent particulars is both intrinsically mathematical in character, and basic to any 'scientific' thought. In the course of a discussion setting out a 'simpler form' of a logical calculus than any that had yet been given, Hilbert gives some further suggestions along the same lines. He claims that we need to recognize

> . . . an *Axiom of Thought*, or, as one might say, an *Axiom of the Existence of an Intelligence*, which can be formulated approximately as follows: I have the capability to think things and to denote them through simple signs (a, b, . . ., X, Y, . . .) in such a fully characteristic way that I can always unequivocally recognize them again. My thinking operates with these things in this designation in a certain way according to determinate laws, and I am capable of learning these laws through self-observation, and of describing them completely.[131]

[131] Unpublished lecture notes dating from 1905 - I owe this reference to Michael Hallett.

(I propose to draw a veil over the suggestion that introspection is the appropriate methodology here.) The foundational nature of this 'Axiom of Thought' is revealed, according to Hilbert, in the fact that the conceptual capacity it describes plays an essential role in the systematic development of the laws of logic itself:

> Arithmetic is often considered to be a part of logic, and the traditional fundamental logical notions are usually presupposed when it is a question of establishing a foundation for arithmetic. If we observe attentively, however, we realize that in the traditional exposition of the laws of logic certain fundamental arithmetic notions are already used, for example, the notion of set and, to some extent, also that of number. . . . . that is why a partly simultaneous development of the laws of logic and of arithmetic is required if paradoxes are to be avoided.[132]

The suggestion that finitary mathematics - or perhaps more accurately, the very basic computational ability of which finitary mathematics is a mathematical theory - is no less than logic part of the necessary basis of any 'scientific' thought recurs in many places throughout Hilbert's writings. In Hilbert [1927], in the midst of a defence of proof theory against Brouwer's charge of 'empty formalism', we find the following:

> The formula game Brouwer so deprecates has . . . an important general philosophical significance. For this formula game is carried out according to certain 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 our understanding, to make a protocol of the rules according to which our thinking actually proceeds.* [133]

This formula game, in which the technique of our thinking is expressed, is of course metamathematics - or equivalently, finitary mathematics. Finitary mathematics, then, either consists of, or perhaps is a mathematical model of this closed system of rules in which 'the technique of our thinking' is expressed.

In the following section, I shall take up the task of making these gnomic suggestions clearer and, hopefully, more plausible.[134] Our current topic, though, is whether finitary mathematics has some special *epistemological* status, and if so, whether Hilbert's notion of intuition is important in establishing that it does.

---

[132] Hilbert [1904] p131.

[133] Hilbert [1927] p27, last emphasis mine.

[134] I should say immediately, though, that I do not find Hilbert's suggestion as it stands in the least absurd. As will soon become clear, I think that there is something in these claims of Hilbert's that is important, intuitive (no pun intended), and very possibly true.

Now, it seems to me that the suggestion that emerges most naturally from these passages is roughly the following. When Hilbert says we must accept that

> . . . . the fact that [the objects of finitary mathematics] occur, that they differ from one another and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects. as something that neither can be reduced to anything else nor requires reduction.[135]

what he means is just that the abilities that we deploy in constructing a tally (matching pebbles with sheep, telling one sheep from another, and one pebble from another, and so on) are part of a minimal conceptual repertoire presupposed by the possibility of thought about an objective world. *Surveyability*, in this sense, serves as a guarantee that the relational properties of strings and elements of strings can be ascertained on the basis of at most finitely many pairwise comparisons of individuals. This is surely a sense of 'intuitive' which is distinctively linked to finitary mathematics. No theorem involving continuity essentially, for example, can be said to be intuitive in this sense, although many such theorems are intuitive in the more relaxed sense used in the title of 'Intuitive Geometry'. There is, therefore, some evidence for the existence of a second, and more philosophically controversial usage of 'intuitive' in Hilbert, related to properties peculiar to finitary mathematics.

If it could be shown that, in principle, the consistency of an ideal theory of classical mathematics (analysis, say) could be established by the exercise of this minimal conceptual repertoire alone - for example, by a demonstration that no proof of the system terminated in an occurrence of the string $\ulcorner 0=1 \urcorner$ - then that theory would have been shown to satisfy the very strongest criteria of rational acceptability possible, since its acceptability would have been shown to be implicit in any conception of an objective, mind-independent reality. In particular, any doubts about the *ontology* of the theory - doubts about whether the 'objects' of the theory, according to its intended interpretation, really existed - would be completely undercut, since the credentials of the theory had been established by nothing more than the very conceptual resources upon which the general idea of existence in a mind independent reality rests. Of course, the envisaged proof of consistency would not establish the existence of what we might call the ideology of the theory - the ideal elements, as it might be, appealed to in the standard interpretation. But this is now mathematically

---

[135] Hilbert [1925] p376.

harmless. Since the credentials of the theory have indeed been established, the puzzling intended interpretation can be taken as an heuristic - just as the ideology of Euclidean geometry is a useful heuristic. This heuristic Hilbert might well be prepared to describe as 'intuitive', but only in the first, philosophically uncontroversial sense.

But now, if something like this really is what Hilbert has in mind - and I think that it is - then it seems to me that we should greet with the greatest skepticism any claim that the *particular proofs* of finitary mathematics have any special degree of certainty, clarity or obviousness. This is just obviously wrong, in virtue of such mundane facts as the length, repetitiveness etc. of finitary proofs. And clearly, it is not an implication of Hilbert's position.

More importantly, though, it would also be at best extremely misleading to claim that the *methods* of proof used in finitary mathematics have any special degree of certainty, clarity of obviousness. It is one thing to know that, in deploying certain kinds of conceptual resource, you are deploying resources the reliability of which cannot intelligibly be questioned: it is quite another thing to know that you are doing so correctly. If the shepherd really has correctly constructed a tally of n pebbles, it makes no sense at all to wonder whether there really are n sheep. It is indeed an implication of Hilbert's position that we can make no sense of there being an inconsistency in elementary arithmetic. But of course, the assurance that the tally has been correctly constructed is no better that the assurance that no sheep has been counted twice, or not at all; that no sheep has disappeared, no 'sheep' is a disguised wolf, no pebble has duplicated itself, and so on ad infinitum - to conceive of the tallying procedure misfiring *is* to conceive of something of this kind having taken place. Nevertheless, it is the very obviousness, simplicity, and clarity of the particular, discrete steps in finitary reasoning which renders the construction of an entire proof of any interesting theorem by finitary means so prone to exactly this kind of computational error. The correct use of finitary means is indeed a guarantee against error of the strongest possible kind, and Hilbert's position implies that this is so. But it is quite another matter to claim that the limitation to finitary means ought to increase our confidence in any mathematical result. This seems to me clearly false, and it is not an implication of Hilbert's position.

The trouble is, Hilbert's way of expressing himself on this matter of the clarity and definiteness of finitary mathematics, coupled with his mixed usage of terms such as 'intuitive', is apt to encourage an interpretation which gets his real position exactly the

wrong way round. One root of the trouble lies, I think, in the temptation to see Hilbert's numerals, the sequences of strokes discussed above, as models of finitary mathematics - to see the relation between the singular term 'three' (or "sss0") and the tally-type '///' as one of reference. Since one can indeed perceive tokens of this type, in a perfectly ordinary sense, one can take 'reference' here to be the very same relation as that which obtains in a non-mathematical context between an ordinary proper name of a type, like 'Old Glory', and the physical objects to which the name refers. A naturalistic account of *this* relation is bound to appeal to some causal concepts, to some kind of causal link between the name and the bearers of the name. It is then very hard to avoid the conclusion that Hilbert is attempting to fit our understanding of mathematics into a broadly naturalistic framework by finding naturalistically acceptable analogues of mathematical objects with which we can causally interact, to which we can 'refer' in exactly this familiar sense.

Talk about the 'intuitability' of finitary (or any other) mathematical truths then appears, fatally, to involve some sort of wonderful capacity to read off the appropriate mathematical properties from arrays of physical objects. Kitcher interprets Hilbert in this way, and then has little trouble in making him seem confused.[136] For if the token arrays

/////

ooooo

are to count as arrays of the same numerical type, as they surely must, and if my mathematical intuition is to detect this, then my mathematical intuition is functioning in such a way as to distinguish the genuinely mathematical properties of the arrays from such properties as their volume, their shape, their location in space and time etc. But talk of mathematical intuition as *detecting* this distinction plainly assumes some independent account of what the distinction *is*, and that account must give us arithmetic without any reliance on mathematical intuition.

This closely resembles a very old and familiar criticism of abstraction as an ultimate source of knowledge. It is found in Berkeley's critique of Locke, but also, and more pertinently, in Frege's criticisms of Mill in 'The Foundations of Arithmetic'. But however things are with Locke and Mill, Hilbert is not vulnerable to this kind of criticism, for this

---

[136] See Kitcher [1976], esp. pp110-114.

interpretation really does get his position the wrong way round. Hilbert is not claiming that our 'intuitive' mathematical knowledge, in either or the two senses of 'intuitive' we have isolated, can be abstracted from models of mathematical assertions: rather, our 'intuitive' mathematical knowledge is what enables us to construct the models.

Let me try to explain a little further. It is an undeniable fact that the selection of an iconography for the presentation of a formal theory T can have the consequence that certain mathematical facts about T become more readily accessible to us, via an inspection of some appropriate diagram. An elementary theorem of Fano's finite plane geometry states that there exist as many 'points' as 'lines'.[137] This is not immediately obvious from the axioms of Fano's geometry, but it just leaps out at you as soon as you attempt, in the most natural way, to produce a drawing satisfying those axioms - a drawing, that is, in which the 'points' of the geometry are put into 1-1 correspondence with dots on a piece of paper, and the 'lines' are put into 1-1 correspondence with lines joining these dots. If the axioms are respected, the result is a diagram which contains exactly seven dots and seven lines. Via the 1-1 correspondence we have established, the diagram can then be converted into a proof that Fano's geometry contains exactly as many 'point' as 'lines'. The theorem is intuitive, in the uncontroversial sense of Hilbert's 'intuitive' geometry. There is nothing especially mysterious about this - or at least, there is nothing suggesting a special faculty of the mind trained in upon the detection of mathematical properties. The properties being detected are physical, not mathematical. The imposition of a certain physical structure on the diagram is guided by our *prior* grasp of the mathematical content of the Fano axioms. Given that we have imposed that physical structure, in a way that is natural to us, on the physical object we produce, it is unremarkable that we are then able, by perusing the diagram, to uncover further mathematical properties of the analogous mathematical structure described in Fano's axioms. The diagram enables to see all at once, so to speak, physical analogues of the mathematical objects and properties presented piecemeal in the axioms.

The same point holds in even in the simplest case. The reason that the array

///

//

[137] For a brief and elementary presentation of Fano's geometry, see e.g. Smart, J.R. [1973] pp14 ff.

is a natural candidate to model the mathematical fact that three is greater than two is that we naturally see the upper array as the result of iterating the operation of writing down a stroke three times, whilst we see the lower array as the result of iterating the same operation twice. There are all sorts of other ways to see this array, although perhaps none of them are so natural to us. None of this does anything to explain the grasp of the mathematical fact represented, the ability to bring mathematical concepts to bear on items encountered in experience. On the contrary, it presupposes this very ability. And nothing in either of Hilbert's usages of 'intuition' denies this.

What may incline one to suspect otherwise is a peculiar feature of finitary theories, such as Fano's geometry, which is lost in ideal theories. You can, for example, diagram any identity amongst natural numbers using Hilbert's stroke-numerals, or any other tallying device for that matter, and then verify the identity by visually inspecting the diagram. The process of verification is intuitive in the uncontroversial sense - in the small at least, it will be immediately obvious on inspecting the diagram that *this* stroke corresponds to *that* stroke, that *this* stroke has no analogue in *that* series of strokes, etc. You cannot verify the 'ideal' identity $(\sqrt{2})^2=2$ in any analogous way. Atomic sentences of finitary theories have diagrams in which physical and mathematical objects are put into 1-1 correspondence. With ideal theories, this feature is lost.

It is then very natural to think that Hilbert's remarks on finitary mathematics quoted above - in particular, the remark that 'the fact that [objects of finitary mathematics] occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects, as something that neither can be reduced to anything else nor requires reduction' - commit him to the view that some special certainty attaches to finitary mathematics in virtue of the verifiability of finitary theorems in this 'intuitive' way.

This claim is just wrong, for the reasons mentioned above. Indeed, it is so *obviously* wrong that it cannot, I think, be what Hilbert intends. The deeper sources of the confusion, in my view, lie in the ambiguities induced throughout the theory of syntax by this fact that elementary truths of syntax - finitary mathematics - can be given physical models. 'Symbol' then becomes ambiguous as between an abstract role in arithmetic, a role *modelled* by a numeral, and a (type or token) physical object: 'surveyable' becomes ambiguous as between a *mathematical* property, mathematically defined over finite sequences, and a *physical* property of arrays of (types or tokens of) physical objects: and

'syntactic property' becomes ambiguous as between, roughly, combinatorial properties of finitary mathematical objects, and physical properties of arrays of physical objects. These ambiguities in turn induce a conflation of the two senses of 'intuitive' we have distinguished. With the *mathematical* understanding of the concepts of syntax in mind, we can say that the theory of syntax is 'intuitive', in the sense that it deals exclusively with those mathematical concepts presupposed by any scientific thought whatsoever. With the *physical* understanding of the theory of syntax in mind, we can say that the elementary truths of the theory of syntax are 'intuitive', in the sense that they can be made immediately obvious on inspection of diagrams, graphs, physical models of some kind. Hilbert's view, and in my opinion the only plausible view, is that a special status is bestowed on finitary mathematics in virtue of its intuitive character in the *first*, but not the second sense.

Consequently, finitary and ideal mathematics do not contrast in any consistent or interesting way with respect to the certainty we can have in any particular result, nor does anything in Hilbert's talk of intuition suggest otherwise. The application of finitary means does not contrast in any consistent or interesting way with the application of ideal means with respect to the certainty that attaches to any result of the application of those means; and Hilbert's position does not imply that it does. This is not to say that there is no epistemological gain to be anticipated from a finitary proof of consistency, however. The gain comes in protecting the ideal theory against what has, historically, been the most prominent objection within the mathematical community to the use of unfamiliar ideal methods - the complaint that nothing in the mathematical world corresponds to these unfamiliar objects. On the picture that is now emerging (I hope), it will be apparent that Hilbert's argument against this kind of objection has two stages. First, the special status of finitary mathematics shows that ontological doubts about the objects of finitary mathematics are groundless, in virtue of the special status of finitary mathematics with respect to any thought whatsoever about a mind independent reality. Secondly, a finitary proof of consistency shows that the apparent ontological excess associated with the ideal elements really is only apparent.[138]

Before going on to discuss the better grounds for attributing a special status to finitary mathematics, I want to pause in order to use the discussion of this section to show that one very influential criticism of Hilbert's Programme, originally offered by Poincaré and more

---

[138] Notice that this reinforces the point that the most fundamental aspect of Hilbert's Programme is the attempt to show that ideal elements' are eliminable from proofs of real theorems in ideal theories.

recently revived, in slightly altered form, by Philip Kitcher, is in fact misguided. This is really by way of an aside, but it may do a little more to clarify the nature of the special status of finitary mathematics.

The criticism concerns Hilbert's use of induction in metamathematics, and it is best introduced in Kitcher's version, since it is possible to respond to Kitcher, I believe, without any detailed discussion of certain technical issues which must be addressed in an adequate response to Poincaré.[139] Kitcher is perceptive in noticing one striking difference between Hilbert's notion of intuition and Kant's. For Kant, Kitcher tells us, intuition can only yield particular arithmetical truths, whereas Hilbert clearly thinks that certain *general* finitary truths are intuitive. Kitcher continues:

> Hilbert's emphasis that '1+a=a+1' expresses a finitary proposition can be supported along with his thesis that we know finitary propositions by intuitive means if we claim that it is possible to intuit a general stroke-symbol. The idea would be that we represent to ourselves sign-designs of the form '1 . . . 1' where we take the dots to stand for an indeterminate number of strokes. By surveying these designs we are able to know for certain basic finitary propositions . . . . For example, we can learn that 1+a=a+1 by first exhibiting the design
>
> 1 . . . 1
> 1 . . . 1
>
> and then transforming it into the design
>
> 1 . . . 11
> 11 . . . 1
>
> From primitive general propositions of this kind we may proceed to more complex results.[140]

The position being attributed to Hilbert, then, is one on which we can come to know certain *general* finitary facts by visually inspecting 'arbitrary', schematic models of those facts. Now, the fundamental facts of finitary mathematics can be regarded as those expressible by numerical equations, and one verifies such equations by way of a pairwise comparison of the constituents of the terms of the equations. And of course, if such a comparison is to be possible, the process of pairwise comparison must terminate after a finite number of steps.

As Kitcher observes, there appears to be a dilemma here for Hilbert, the horns of which are formed by the two possible ways of construing these schematic numerals. If we take the

---

139 I return to Poincaré's criticism at greater length below.
140 Kitcher [1976] p110.

schematic numeral intuited to have a definite number of strokes, then we get termination all right, but we get a particular rather than a general fact - the fact that decomposition of an n-element stroke symbol terminates for some *particular* value of n. On the other hand, if we take the arbitrary stroke symbol to have an indeterminate number of strokes, then we have some prospect of getting generality, but no assurance that the decomposition procedure will terminate. That assurance requires either a unique element, or no element at all, to the right (say) of any given element of the array, and this is not guaranteed by an intuition in which the intuited stroke symbol has an indeterminate number of strokes. In short: Hilbert seems to be saddled with all the unattractive aspects of Locke's notorious 'arbitrary triangle'.

This is, I think, a very serious objection to one way in which the idea of intuition might be used in a philosophy of mathematics. But it is not an objection to Hilbert's usage, for at bottom it rests upon just those ambiguities between physicalist and mathematical readings of the key syntactic notions, and the resultant confusion of two senses of 'intuition', that we have been discussing. In 'surveying' a particular diagram, or proof, the 'syntactic' properties I am attending to are physical properties of a particular physical object. Since this object has been constructed out of its constituents with an eye to producing clearly detectable physical analogues of the notions of syntax, in its mathematical sense, I will be often able to recover a good many syntactic properties of the mathematical structure being modelled by visual inspection of the physical model.

More than this, though, I will be able to bring to bear my knowledge of the procedure according to which the physical model has been constructed - the knowledge which enables me to produce physical models of a great many different particular mathematical structures. If I know how to model the finitary identity '2+2=4', I will also know how to model the identity 'a+b=n' for at least a great many particular values of a, b, and n - one constructs a tally of a, a tally of b, a tally of a+b, a tally of n, and then proceeds to a pairwise comparison with respect to 'length' of the last two tallies constructed. The assurance I have that this last operation will terminate is provided, not by any amount of physical inspection of particular tallies, but rather by my grasp of the construction routine for tallies: a tally is the result of finitely many applications of the basic procedure of adding a unit to a tally. And this assurance is 'intuitive', not in the sense in which it can be verified by the inspection of models, for it cannot, but in the sense that assurance is provided by the resources of this minimal mathematics - finitary mathematics - alone.

The likely response to all this, once more familiar from the debate over Locke's arbitrary triangle, is this. If what is known intuitively is a rule for constructing integers, there is simply no need for the further intuition of particular integers. In general, if one has grasped a rule for applying a concept, one has no further need for a paradigm case of the application of the concept, which is presumably what an intuition of an object satisfying the concept is supposed to be. Indeed, it is precisely the ability to apply the rule governing the concept that enables one to identify a paradigm case as a paradigm case of the application of *that* concept. Given the intuitive grasp of the rule, then, intuition of objects is not needed.

But this is not an objection to Hilbert: it is pretty much Hilbert's own view. Consider the following passage, in which Hilbert is responding to Poincare's criticism of Hilbert [1904]:

> . . . he [Poincare] denied from the outset the possibility of a consistency proof for the arithmetic axioms, maintaining that the consistency of the method of mathematical induction could never be proved except through the inductive method itself. But, as my theory shows, two distinct methods that proceed recursively come into play when the foundations of arithmetic are established, namely, on the one hand, *the intuitive construction of the integer as numeral (to which there also corresponds, in reverse, the decomposition of any given numeral, or the decomposition of any concretely given array constructed just as a numeral is), that is, contentual induction, and, on the other hand, formal induction proper, which is based on the induction axiom and through which alone the mathematical variable can begin to play its role in the formal system.*
> 141

'*The intuitive construction of the integer as numeral*' - that is, the recursive procedure for generating numerals, or, as I have been using these notions, the abacus construction of abacus configurations. Exactly what this distinction between kinds of induction comes to will require much further discussion: the point here, though, is that it is this 'intuitive construction of the integer as numeral', and not any 'intuitive' apprehension of arbitrary stroke-symbols, that is central to Hilbert's account of our knowledge of general facts in finitary mathematics.

But it is a mistake to conclude from this that the 'intuitive' inspection of figures (in the alternative, physical sense) is redundant. Consider the task of programming a computer to discover theorems in a simple finite geometry, for example. One might try a brute search algorithm: just set the machine up to churn its way through an enumeration of the deductive consequences of the axioms of the geometry, time without end. This gives lots of theorems, but on the other hand, we may have to wait a long time for the theorems

---

141 Hilbert [1927] pp472-473, my emphasis.

which we would be interested in - the ones which give us some insight into the system we are exploring, for example. On the other hand, one might program the machine to construct diagrams of the axioms in very much the same way we do, with dots and lines, say, and then provide an algorithm that recovered theorems from a 'survey' of those diagrams. This might involve sacrificing the assurance that all the theorems we are interested in will, sooner or later, be discovered by the machine, but on the other hand, it may also lead to the consistent generation of theorems that strike us as interesting. In doing mathematics, we do not just flounder around in theorems: we select, we look for interesting patterns - interesting, that is, according to our lights, according to the mathematics that we already understand. This is where intuition, in its other sense, plays its part - not as a warrant for truth, but rather as a guide to significance. The paradigm, the particular object constructed according to the rules of the system, guides our application of the rules, not in the sense of showing us *how* to apply the rules - if we did not already know that, we would not know that it was a paradigm - but rather in the sense of showing us *where* to apply them. The paradigm makes our attention to the deductive consequences of an axiom set selective.

What now have to hand, then, are a good many reasons to reject options (A) and (B) - good reasons, that is, to view any claims about the special ontological status, or the special epistemological status of finitary mathematics with skepticism. Finitary mathematics is not 'about' expressions, and its truths are not, in general, particularly obvious. It remains to consider option (C).

Section Four: Finitary Mathematics and Mathematical Intuition. Option (C), you will recall, was suggested by the following passage from Hilbert [1925]

> Kant already taught - and indeed it is part and parcel of his doctrine - that mathematics has at its disposal a content secured independently of all logic and hence can never be provided with a foundation by means of logic alone; that is why the efforts of Frege and Dedekind were bound to fail. Rather, as a condition for the use of logical inferences and the performance of logical operations, something must already be given to our faculty of representation [in der Vorstellung], certain extralogical concrete objects that are intuitively [anschaulich] present as immediate experience prior to all thought. If logical inference is to be reliable, it must be possible to survey these objects completely in all their parts, and the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects, as something that neither can be reduced to anything else nor requires reduction. This is the basic philosophical position that I consider requisite for mathematics and, in general, for all scientific thinking, understanding, and communication. And in mathematics, in particular,

> what we consider is the concrete signs themselves, whose shape, according to the conception we have adopted, is immediately clear and recognizable.[142]

Having devoted some energy to debunking notions of mathematical intuition, I now intend to introduce and explain a different notion of mathematical intuition, suggested by this passage, and relate it to the discussion of mathematical objects of the last two sections of **Chapter One**. The notion in question is also present, in my opinion, in Gödel's philosophical writings - unsurprisingly, since both Hilbert and Gödel are drawing upon a common understanding of the teachings of Kant. I do not intend to take up the topic of Kant's conception of mathematical intuition, and I make no claims as to the Kantian credentials of the Hilbert/Gödel notion. But I do think that it will be helpful to spend some time discussing Gödel's views.

Gödels' main discussion of mathematical intuition is to be found in a philosophical article he devoted to Cantor's continuum problem, which exists in two versions. The earlier dates from 1947, before it was known that the continuum hypothesis (**CH**) was independent of set theory. The article was then revised and reprinted in 1964 for the Benacerraf/Putnam anthology on the philosophy of mathematics, together with a remarkable four-page supplement. By this time, Gödel knew that (**CH**) was independent of set theory. The following passage occurs in the 1964 supplement. Gödel has been expressing his conviction that, despite its independence of the axioms of **ZF**, (**CH**) must have a definite truth value, and this has led him into a discussion of the grounding of the concept of set. He writes:

> . . . despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see any reason why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception, which induces us to build up physical theories and to expect that future sense perceptions will agree with them, and, moreover, to believe that a question not decidable now has meaning and may be decided in the future. . . .
>
> It should be noted that mathematical intuition need not be conceived of as a faculty giving an *immediate* knowledge of the objects concerned. Rather it seems that, as in the case of physical experience, we *form* our ideas also of those objects on the basis of something else which is immediately given. Only this something else here is *not*, or not primarily, the sensations. That something besides the sensations actually is immediately given follows (independently of mathematics) from the fact that even our ideas referring to physical objects contain constituents qualitatively different from sensations or mere combinations of sensations, e.g., the idea of object itself, whereas, on the other hand, by our thinking we cannot create any qualitatively new elements, but only reproduce and

---

142 Hilbert [op cit] p376.

> combine those that are given. Evidently the 'given' underlying mathematics is closely related to the abstract elements contained in our empirical ideas.* It by no means follows, however, that the data of this second kind, because they cannot be associated with actions of certain things upon our sense organs, are something purely subjective, as Kant asserted. Rather they, too, may represent an aspect of objective reality, but, as opposed to the sensations, their presence in us may be due to another kind of relationship between ourselves and reality.
>
> However, the question of the objective existence of the objects of mathematical intuition (which, incidentally, is an exact replica of the question of the objective existence of the outer world) is not decisive for the problem under discussion here [the problem of (CH), that is]. The mere psychological fact of the existence of an intuition which is sufficiently clear to produce the axioms of set theory and an open series of extensions of them suffices to give meaning to the question of the truth or falsity of propositions like Cantor's continuum hypothesis. What, however, perhaps more than anything else, justifies the acceptance of this criterion of truth in set theory is the fact that continued appeals to mathematical intuition are necessary not only for obtaining unambiguous answers to the questions of transfinite set theory, but also for the solution of the problems of finitary number theory (of the type of Goldbach's conjecture), where the meaningfulness and unambiguity of the concepts entering into them can hardly be doubted. This follows from the fact that for every axiomatic system there are infinitely many undecidable propositions of this type.
> (* Note that there is a close relationship between the concept of set explained in footnote 14 [this is a version of the 'iterative' concept of set] and the categories of pure understanding in Kant's sense. Namely, the function of both is 'synthesis', i.e. the generating of unities out of manifolds (e.g., in Kant, of the idea of *one* object out of its various aspects.) [143]

Now as I understand it, there are *two* notions of intuition under discussion in this passage. In the last paragraph, Gödel claims that the 'mere psychological fact of an intuition which is sufficiently clear to produce the axioms of set theory and an open series of extensions of them suffices to give meaning to the question of the truth or falsity of propositions like Cantor's continuum hypothesis.' I take it that the claim here is simply that there is, as a matter of psychological fact, sufficient consensus amongst the mathematical community as to what is and what is not immediately apparent in set theory to permit eventual agreement on (CH). What is intuitive, in *this* sense, is just what is found obvious. Gödel's claim is that our intuitions in this psychological sense, perhaps guided by informally presented 'genetic' concept of set outlined by Zermelo, are sufficiently strong and clear to permit agreement on axioms beyond those of standard ZF which will suffice to settle the (CH) to our mutual satisfaction. Now, staying at the level of empirical psychological fact, I think that this is actually false. From what I can tell of the mathematical community, and especially the set theorists, from the outside, it appears to me that no such consensus exists. But this, I think, is philosophically unproblematic and without implications for the ontology of mathematics. For the activity that Gödel is appealing to here - in effect, the

---

[143] Gödel [1964] p268.

investigation of models of set theory - is an exercise in interpretation only in the first of the two senses distinguished in **Chapter One**. It is perfectly possible to join in the discussion of whether or not this or that model of set theory, in which (CH) holds or fails as the case might be, keeps faith with our intuitive understanding of the concept of set in the psychological sense, whilst remaining a strict formalist at the level of ontology.

But the earlier parts of this quotation seem to me to feature a conception of mathematical intuition which is neither psychological in character nor ontologically innocent. It is this notion that figures in Gödel's claims about the 'given' underlying mathematics, and it is this notion that is said to bear a close relationship to the Kantian categories of pure understanding. This is the notion that we need to understand.

Let us begin by setting aside some misguided criticisms that have appeared in the literature on this passage in Gödel. Crispin Wright complains that Gödel . . .

> . . . postulated a special intuitive faculty, akin to a kind of perception of mathematical objects, to explain our capacity to know mathematical truths. Such a postulation, of course, *explains* nothing of the sort. The picture, indeed, threatens to push our recognition of the truth of a mathematical statement beyond philosophical account. [144]

And to be sure, we have just seen that Gödel does indeed speak of his faculty of intuition as 'something like' a perceptual capacity. But alike in what respect?

Now, in the ordinary way, we talk of perception of *objects*, and perception of *truths* - of perception de re, and perception de dicto, one might say. Is the claim that mathematical intuition yields perception de re of sets? That is certainly a very suspicious notion. But then, at the beginning of the second paragraph, Gödel explicitly denies that his mathematical intuition gives immediate - i.e. de re - knowledge of mathematical objects. So the analogy with perception is more likely to be with perception de dicto, or perception of truths.

Now, why exactly is the postulation of a special quasi-perceptual capacity in this de dicto sense said to be unexplanatory? Well, we can certainly agree that an explanation that rests upon this kind of postulation is worthless if nothing whatsoever is known or knowable of the alleged special faculty other than its manifest effects. This is the defect famously

---

[144] Wright [1980] p3.

satirized in Moliere's 'explanation' of the soporific powers of opium in terms of the possession of a 'virtus dormativa' - as we might put it, a sleep-inducing disposition. Wright's complaint, then, may be that Gödel's postulation of a faculty of mathematical intuition is at bottom no better than the postulation of a basic disposition to know some mathematical truths.

But this kind of defect is alleviated if some characterization of the special faculty can be given in terms relatively independent of the data for which an explanation is sought. In the case of the sleep-inducing properties of opium, for example, postulation of a virtus dormativa may become genuinely explanatory if some characterization of the virtus dormativa (and of the subject upon which the virtus dormativa works) can be given in *chemical* terms. The result might be an explanation of the sleep-inducing disposition of opium at the level of basic chemistry. Drowsiness could perhaps be shown to be associated with such and such chemical changes in the brain, and opium could be shown to have a chemical constitution apt to cause those chemical changes. The question we must now ask is, Why exactly does this kind of move help?

To begin with, notice that it would not be plausible to claim that the postulation of a special mental faculty is legitimate only if the alleged faculty can be finally characterized in non-dispositional terms. For, to stick with our opium example, the characterization of the sleep-inducing disposition of opium in chemical terms may very well itself be dispositional: after all, the primitive expressions of chemistry, and indeed of basic physics, appear to be prime candidates for dispositional characterization. Taking science as one finds it, it is simply untrue that scientific explanations couched in dispositional terms may always be regarded as placeholders for explanations couched in non-dispositional terms.

Nor would it be plausible, in my view, to press any general demand that explanations at whatever level must shown to be reducible, even in principle, to explanations couched in the vocabulary of a non-mentalistic science, leave alone to explanation in the vocabulary of basic physics. Any such demand, if taken fully seriously, would deny almost all of cognitive science as it is currently practised any genuine explanatory value. Perhaps that conclusion would not deter some, but it seems to me that, if the best that can be done against Gödels' postulation of a special faculty of mathematical intuition is an argument that would close down the psychology departments, then Gödel is home free. So the point had better not be that Gödel's talk of mathematical intuition is hopelessly unexplanatory because

it is either irreducibly dispositional, or cannot be cashed out in terms of some more basic, non-mentalistic science.

But surely, you will say, there is no mystery about why the attribution of a virtus dormativa to opium cannot explain why opium causes drowsiness. For such an 'explanation' is no different from the bald assertion that opium *just does* cause drowsiness - we are given no information whatsoever about *why* opium causes drowsiness. And of course that is right. What we need to guard against, though, is an unduly restrictive conception of what it takes for talk of the sleep-inducing properties of opium to become genuinely explanatory. The temptation, of course, is to think that what is invariably required for a genuine explanation is an account of a causal mechanism, such as that linking ingestion of opium and drowsiness, with this talk of causal mechanisms then cashed out in turn in terms of causal links between types of states individuated in the vocabulary of physics. This conception really is far too restrictive, for the reasons I have just mentioned.

A better line of thought, I think, begins with the observation that an account of the virtus dormativa of opium in chemical terms enables us to subsume the explanation of the sleep-inducing properties of opium within a body of explanatory theory of much broader scope - a comprehensive theory which explains, say, the sleep-inducing properties of other kinds of substance, along with a wide variety of related psychological and somatic effects of a wide variety of narcotics. To see the most important point here, suppose that we know only that opium has a sleep-inducing disposition, and that coffee, say, has a bowel-moving disposition. What happens if we put some opium in our coffee? Does the resultant substance have both dispositions, does one prevail over the other, do they cancel each other out? No answer is possible within the confines of a theory which provides only a *piecemeal* attribution of dispositions. However, we would expect an adequate theory, pitched at the right explanatory level, to provide a prediction covering exactly this kind of combination of dispositionally characterized properties. Now of course, unpacking this talk of the 'right' explanatory level is itself a complicated business, but we surely do have some intuitive grasp, prior to the elaboration of any scientific theory, of some range of related phenomena of which the sought for theory ought to provide unified explanations.

In the case of Gödel's talk of mathematical intuition, the relevant theory must be an overall theory of our cognitive capacities, both empirical and non-empirical. But is it then *obvious*, from what we see of Gödel's talk of a faculty of mathematical intuition in the passage above, that no integrative account of the functioning of mathematical intuition

within such an overall theory of our cognitive capacities is possible? It does not seem at all obvious to me. To be sure, we will want to know much more about the workings of this mathematical intuition, but I see no reason to believe that nothing more can be possibly be said on that topic. One might indeed conclude otherwise *if* Gödel was postulating the existence of a faculty that gave us some kind of direct, unmediated access to Cantor's paradise - perception de re of sets. But that is not in fact Gödel's position, as he clearly states.

Beyond these relatively specific ways in which Wright gets Gödel's position wrong, though, Wright's complaint seems to me to betray a more general misunderstanding of what Gödel is trying to do. Wright hears Gödel's notion of mathematical intuition as intended primarily to address worries about the special conditions required for mathematical knowledge - in effect, to supply us with an analogue of causation, of a reliable, knowledge-yielding mechanism to relate us to mathematical truth. But I think that it is at least equally reasonable to hear Gödel's notion as in fact responding to a second kind of difficulty, concerning the conditions under which beliefs about the characteristic objects of transfinite set theory can be ascribed to finite cognitive systems - a problem which is fundamental to Hilbert's Programme, as I have described it. And if we do hear him in this way, then the talk about the given' underlying mathematics, about the 'abstract elements' in our thought about the empirical world, will appear in a somewhat different light. But I think it best to approach this interpretation slowly, by way of some further criticisms of Gödel offered by Charles Chihara.

Recall that Gödel claims that mathematical intuition shows itself in the phenomenon he describes as the axioms of set theory 'forcing themselves upon us as true'. Chihara, in his criticisms of Gödel, finds this a thoroughly occult notion.[145] He invites us to consider the following case. A student takes up Schoenfield's text on set theory and reads the following passage, in which Schoenfield explains the 'iterative' concept of set mentioned by Gödel in our last quoted passage:

> We start with certain objects which are not sets and do not involve sets in their construction. We call these objects *urelements*. We then form sets in successive stages. At each stage we have available the urlements and the sets formed at earlier stages; and we form into sets all collections of these objects. A collection is to be a set only if it is formed at some stage in this construction.[146]

---

[145] See Chihara [1982], and Chihara [1990] pp15-20.
[146] Schoenfield [1967] p238.

A little later, our student comes across the regularity axiom of set theory, which tells us that, if a set S has an element at all, then it has a minimal element - an element y, that is, such that S and y are disjoint. Reflecting on what she was told above about what sets are, our student might reason as follows. Suppose S has members. Let y be a member of S formed at as early a stage as possible in the construction procedure described above. Now, if y in turn has any members, they must have been formed at an earlier stage than y, so they cannot be members of S. So S and y must be disjoint, and the regularity axiom must be true.

Now, I agree with Chihara that, if this is all that is meant by saying that the axioms of set theory force themselves on us as being true, then there is absolutely no reason to think that this experience provides evidence for the objective truth of the regularity axiom, or any other part of set theory, for two reasons. Firstly, nothing even remotely analogous to a special kind of perception is involved here - simply the ability to reason deductively. And secondly, the kind of interpretation of the axioms of set theory that is involved here is the first, thin sense distinguished in **Chapter One**. Nothing in the interpretation of set theory in this sense can induce an ontological commitment to sets. But the earlier parts of the passage quoted above makes it clear that this is *not* the only kind of experience Gödel has in mind when he speaks of the axioms of set theory as forcing themselves upon us as true.

However, Chihara raises a much more serious worry. On one possible reading, Gödel is claiming that this phenomenon of convergence in mathematical enquiries is to be explained via the postulation of a shared mathematical intuition. And this can seem a suspicious idea, for the following reason. In the natural sciences, the phenomenon of convergence on the truth is standardly explained in *causal* terms. The scientist forms hypotheses, and tests them by experiment. Mother Nature either shows the hypothesis to be false, or allows it to stand as partially confirmed. In this way, with much skill and ingenuity, the range of initially plausible competing explanations can be narrowed, and the narrowing process - convergence - is governed ultimately by the intersubjectively available evidence of the senses. Plainly, this story is causal in nature. And equally plainly, convergence in mathematics cannot be explained in terms of set-theoretic reality impressing itself upon us in *this* sense. If mathematical intuition is intended to supply us with a surrogate for causation in order to enable us to import this kind of naturalistic explanation of convergence over to the mathematical case - and this, I think, is what motivates Chihara's complaints - then Gödel is indeed intruding upon the proper explanatory domain of natural science in a

way we ought to be reluctant to accept. As Chihara points out, not only is this story of dubious intelligibility, but there is also the possibility of a far more plausible naturalistic alternative. For there seems to be some hope that cognitive psychology might explain this phenomenon of convergence by uncovering some shared cognitive apparatus that mediates our ability to reason mathematically.

Now Chihara's suggestion is not, of course, that mathematical truth might *itself* have an explanation in terms of psychological facts about mathematicians. An empirical explanation is envisaged here for *convergence* on mathematical truth, not for the mathematical truth converged upon. Still, even this does not appear to be quite right, since convergence on at least basic mathematical truths is surely a constitutive feature of rationality, and thus something that *must* characterize rational agents, rather than something we might discover about rational agents.

But this suggests that there are in fact two questions here, which need to be distinguished carefully. Chihara's interest is in the question of how the natural world can come to contain such things as mathematicians, and he thinks, as I do, that this is something that might have a naturalistic explanation of a familiar empirical kind. I see no reason to think that Gödel need deny this. The other question, though, concerns rather the issue of what the cognitive capacities of rational agents must be like, given the constitutive truth that they must be capable of reaching convergence upon at least the most fundamental mathematical truths, as well as the simplest truths about their shared environment. This is *not* an empirical question (although an answer to it will incur empirical liabilities), and an attempt at explanation here need not be in competition with any explanatory task proper to natural science. And it seems to plausible to read Gödel as claiming that mathematical intuition, as he understands it, must have a role to play in such an account.

What is in question, on this reading of Gödel, is not an empirical enquiry into the grounds of convergence as Chihara understands it, but rather an a priori investigation of what cognitive capacities we must have, given that our basic mathematical and non-mathematical knowledge is roughly as we take it to be. There is nothing exotic or occult about this kind of suggestion, in my view. And it points us in the right direction, if we want to understand Gödel, and Hilbert, for it leads towards issues concerning the ascription of mathematical concepts.

By now, two things will have occurred to you. Firstly, we have been hearing a lot about what Gödel's conception is *not*, but very little about what it *is*. Secondly, Gödel appears to have been saved from his critics by draining his doctrine of all interest, and in particular, by ignoring this problematic analogy of mathematical intuition with perception. It is time to try to remedy these defects.

A sympathetic attempt to understand Gödel must begin, I think, with the parallel he suggests between sets and objects, or rather, between the ability to think of sets, and the ability to think of objects. Gödel's claim is that we form our ideas of physical objects, or - better - the idea of a physical object in general, on the basis of sensations, which alone are 'immediately given'; although the idea of a physical object contains elements 'qualitatively different' from sensation or collections of sensations. And in some analogous way, Gödel suggests, we form our idea of set in general out of something else which is immediately given - although *not*, in this case, something immediately given in sensation. In the important footnote to this passage, the suggestion is made that formation of the set concept from something immediately given is a special case of the general Kantian notion of synthesis, the activity of the mind which unifies the chaos of sensory impressions *prior* to, and as a precondition for, the conceptualizing activity of the understanding.

What is meant by 'object' here? There is a very thin notion of object, which I can best explain by reminding you of one aspect of Russell's philosophy in his logical atomist period. You will recall that, in the logically perfect language Russell was then seeking, there was to be a class of expressions Russell calls genuine singular terms. Each genuine singular term had as its reference a sense datum - a little patch of color, a momentary whiff of scent, the pointilliste bric a brac of a Buddhist metaphysic. The question of unperceived existence cannot even be raised for objects in this thin sense. To understand Gödel, you must not think of objects in this way. Rather, you will have to think of objects - or physical objects, as I shall sometimes say - as neither bundles of sense data, nor permanent possibilities of sensation, but as the thick, full-blooded, mind- and experience-independent particulars of common sense and physical theory as realistically construed.

It is an ancient commonplace of philosophy in the empiricism tradition, apparently echoed in Gödel's remarks and in the Quinean doctrines that dominated the closing sections of **Chapter One**, that the notion of a physical object in this thick sense is underdetermined by what is 'immediately given' to us in experience, i.e. by experience of objects in the thin sense. As we saw, in the Quinean response to this gap we are said to 'posit' physical

objects, along with mathematical objects and such things as the gods of Homer, in a spontaneous outpouring of theory designed to render cognitively tractable the flux of experience. In the progress of scienc , and at the continued promptings of experience, theory becomes somewhat less spontaneous, and the Homeric gods pass away in favor of simpler, more comprehensive, more testable hypotheses. To date, however, physical objects and the objects of mathematics remain, as commitments of physics - the best theory of the world we have.

It seems to me that Gödel can accept all this. This can be part of his story about why we should believe there are mathematical objects. It is a story in which mathematical intuition appears to play no part. But we should then ask: What is meant by 'positing' objects? Why does the torrent of experience prompt in us the spontaneous theories it does, rather than any of the myriad alternatives? As I read him, Gödel and Hilbert alike are offering a Kantian kind of answer to questions of this sort. Let me now part company with Gödel and Hilbert, and tell you how I think this story must go.

For Kant, on my best understanding of him, objects in the thick sense are a precondition of experience. I would expand this thought somewhat as follows (and I am aware that I depart here from Kant). If a creature has a behavioral repertoire complex enough to resist characterization simply in terms of stimulus and response - that is to say, a repertoire that cannot be characterized adequately without appeal to some notion of mental representation - then any theory attempting to explain the cognitive capacities of the creature will have to draw upon the notion of a physical object. Roughly, the theory will have to locate the creature in an environment of physical objects. Notice that what is in question here is not the concepts that we must attribute to a creature capable of representing its environment: what is in question is the weakest set of assumptions one can make about the environment of a creature in order for the idea of concept-deployment to get a grip.

Indeed, I would want to go a little further than this, and claim that it is a further presupposition of representational theories that concept-deploying creatures must be located in environments of discrete, distinguishable objects standing in relations, to each other and to the creature, of the type describable by phrases such as 'to the left of', 'behind', 'above', 'further from', and so on. Objects standing in arrays governed by such relations, I believe, instantiate initial segments of the simplest mathematical structure, that of a well-founded linear order - the order instantiated in the natural number sequence. In my opinion, arrays of objects *ordered in this way* are the best candidate for what Gödel calls

the 'given' underlying mathematics. On this view, then, what is given to us in mathematical intuition is immediate experience of finite segments of the well-founded linear order type. The claim is that it is a non-empirical presupposition of concept deployment that experience be experience *of* objects standing in these structural relations. Any theory of representation must presuppose objects instantiating this structure.

It may be helpful here to think for a moment of how you were first introduced to Turing machines and the kinds of problem they are capable of solving. The details of the exposition vary from text to text, but you were certainly introduced to some 'conventions' governing the structure - the *physical environment* - cognized by the machine. In one standard variant, the environment consists of a *tape*, divided into *squares*, containing certain *symbol tokens*. The machine then operates by moving to the *left* or *right*, one square at a time, in a series of *discrete* jumps. Some configuration will be designated as the *halting configuration*, leaving the value of the function computed in a fixed position on the tape determined by the input and the machine's instructions. All of this is designed to ensure that a certain structure - that of a well founded linear order - is instantiated in the physical environment of the machine, in a form that the machine can 'recognize'. There are various ways to achieve this, and in this sense the particular ways selected are indeed conventional. What is not a matter of convention, though, is the necessity to ensure that these structural conditions on the environment of the machine are satisfied. Of course, all of this explanatory activity is by way of making vivid a series of conditions that can be given a purely mathematical interpretation. Turing machines, after all, are *functions*. The theory of computability can be interpreted in the weak sense - modelled in a weak set theory, for example - and when interpreted in this way, it is a part of pure mathematics, innocent of ontological commitment.

This mention of Turing machines obliges me to break off from the main line of discussion, in order to say few words on the identification of finitary mathematics with PRA. In Tait [1981] this identification is defended in detail, and I have nothing of substance to add to his arguments.[147] Notice, though, that the Turing machine has capacities that considerably

---

[147] In particular, I take the same view as Tait on the well known passage in Hilbert [1925] which has been taken by some (see Kreisel [1970], for example) to license a far more liberal identification of finitary mathematics - see Tait [1981] pp544-545 and refs. therein. The succinct discussion of this point in Simpson [1988] is also worth consulting.

The kind of liberal identification of finitary mathematics advocated by Kreisel, however, does have some echoes in Hilbert's thought, especially towards the very end of his active life, when he learned of the incompleteness theorems. Hilbert was convinced that those theorems did not preclude the possibility of a finitary consistency proof for arithmetic, and there is perhaps some grounds for suspicion that he was

exceed the bounds of the primitive recursive. In particular, functions defined by the minimization (or least search) operator are Turing computable, but do not belong to PRA. There are several points to be noticed about this.

Firstly, the least search operator is closely akin to the Hilbert ε-symbol discussed in **Chapter One** - the exemplar chosen by the ε-symbol can always be taken to be the least such. As we saw, Hilbert accounted the ε-symbol the sole ideal element in his preferred formulation of arithmetic, and this supports the contention that it ought not to count as finitary. Secondly, it is well known that the least search operator introduces into the theory of computability operations which are not guaranteed to halt (if no least φ exists), and this destroys the effectiveness that is an essential feature of the finitary. Thirdly, the least search operator facilitates the computation of functions that Hilbert certainly did not count as finitary. The standard example is the Ackermann function defined by

$$\text{Ack}(0, n) = n + 1$$
$$\text{Ack}(m+1, 0) = \text{Ack}(m, 1)$$
$$\text{Ack}(m + 1, n + 1) = \text{Ack}(m, \text{Ack}(m + 1, n))$$

This function is computable, but not primitive recursive.[148] In particular, it grows faster than any primitive recursive function. Indeed, it is a useful exercise to calculate a few values of this function - up to Ack (5, 5), say - since doing so is guaranteed to make you love the method of ideal elements.[149]

Thus, whilst it is easily seen that Ack(m, n) is defined for all values of m, n, the proof essentially involves a double induction which is not plausibly regarded as finitary, and which was not so regarded by Hilbert.

---

tempted here towards allowing some forms of transfinite induction to count as finitary - enough, perhaps, to enable him to accept Genzen's 1936 consistency proof. But this is speculation. Bernays, of course, went on to pursue a liberalized Hilbert's Programme, with an extremely generous notion of the finitary. This tradition continues in the work of Kreisel and Feferman on predicative systems of analysis (see e.g. Feferman [1964]). For the relation of this work to the original Hilbert Programme, I take the view of Simpson [1988]. In general, I think that philosophical (as opposed to mathematical) interest declines very rapidly as the bounds of the finitary are loosened.

[148] See Hilbert [1925], pp388-389. For a very detailed discussion of the Ackermann function, see Yasuhara [1971], and for a concise and readily accessible account, see Epstein and Carnielli [1990] pp110-115.

[149] For a suggestive and stimulating philosophical discussion of the Ackermann function, see Boolos [1987].

This is perhaps the best place to mention Hilbert's final point in his defence of his work against the criticism of Poincaré.[150] In response to Hilbert [1904], Poincaré pointed out that classical mathematics makes very extensive use of the principle of induction. But the only possible way of giving a syntactic proof of consistency for a formalized mathematical theory was by induction. Hilbert was therefore trapped in the circle of using induction to justify the use of induction. The response to this is obvious from the discussion of the Ackermann function: there is induction and induction. And in particular, there is no circularity in using induction restricted to formulas defined by primitive recursion, say, in metamathematical reasoning about PA (for example), in which induction is available for a more extensive class of formulas. This is the kind of distinction that Hilbert has in mind in Hilbert [1927] pp472-3, where he speaks of induction 'based on the intuitive construction of the integer as numeral' as opposed to induction which makes unrestricted use of the induction axiom.

Let us return now to the main line of argument. If the fundamental thesis of cognitive psychology is correct - the thesis, that is, that the mind is essentially a kind of complex digital computer (or complex of digital computers), that cognitive processes are essentially computational - then the presence of this structure in the physical environment of any system the behavior of which demands explanation in cognitive terms becomes *a commitment of natural science*. The natural number sequence, thought of as potentially infinite in the manner of the Turing machine's tape, is a presupposition of the representational theory of the mind.

Recall now the crucial analogy between mathematical intuition and perception. If anything has become common ground amongst theorists of perception, whether scientific or philosophical, it is the view that any theory of perception must assume a relatively rich 'innate' cognitive endowment which is brought to bear immediately, without explicit learning or inference, on the items encountered in perceptual experience. To oversimplify, we might say that it must count as a basic, not a derived fact of our perceptual experience, that oranges are seen as subjectively more similar to apples than to bananas in shape (or whatever) - that we associate oranges and apples, but not oranges and bananas, under the sortal concept 'round'. In fact, though, talk of innateness is misleading here, for no empiricist need deny that *some* such basic similarity space has to be assumed if perception

---

[150] See Poincaré [1908], pp169-171. The same complaint is found later in Brouwer [1912] p71.

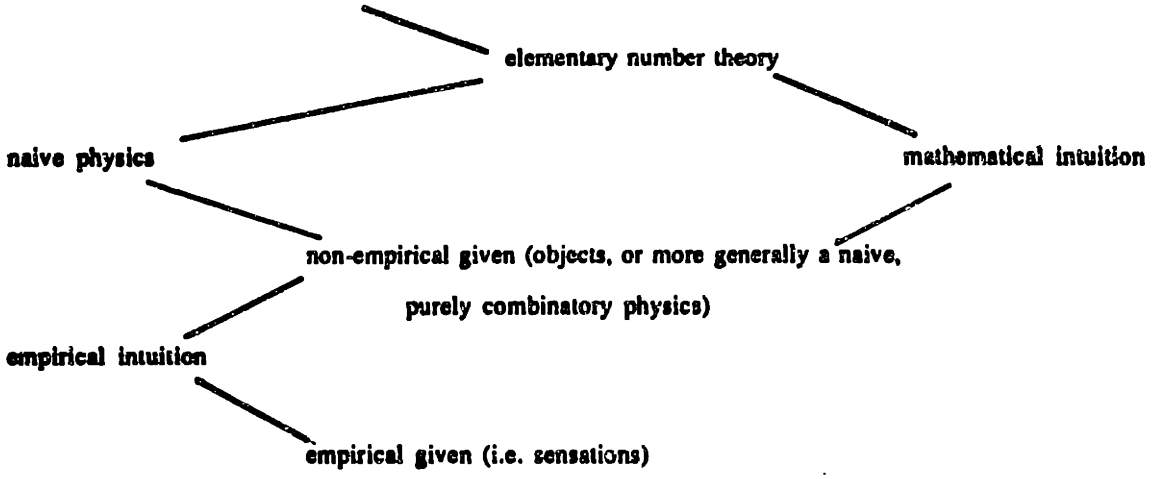
is to be possible. The point is really a point about the weakest assumptions that must be made, if perception of an objective world is to be theoretically intelligible at all.

The basic contours of any similarity space, it seems to me, can reasonably be described as intuitive. That concept deploying creatures must have *some kind* of perceptual intuition, must immediately deploy similarity judgments, is not a claim that belongs to empirical psychology, however. Equally, the claim that concept-deployment presupposes some kind of environment of objects instantiating initial fragments of a well-founded linear order does not belong to empirical psychology.

But what does this have to do with sets? The Gödelian claim, remember, is that the concept of set plays the same role with respect to the given underlying mathematics - which I have now associated with initial fragments of a well-founded linear order - as the concept of physical object plays with respect to what is given in sensation.[151] The Gödelean thought seems to be this: just as we construct ever more sophisticated physical theories from the extremely primitive, fragmentary physics that is given together with the concept of object, the better to manage the ongoing flux of experience, so too we construct ever more complex set theories (perhaps not under that description - here we think of the theories of the integers, the reals, the complex numbers, as fragments of set theory) from the extremely primitive set theory that is given together with the well-founded linear order type, the better to understand the mathematical properties of the potentially infinite natural number sequence. The picture, then, is somewhat as follows:

---

151 It is worth reflecting on what this shows about Gödel's intentions with respect to mathematical intuition. For it seems clear to me that the doctrine he advances answers in the first instance to a worry *about what is required for the basic mastery of the concept of set*. Appeals to intuition have always seemed suspicious on the grounds of dogmatism and arbitrariness, for what is to prevent anyone from seeking to protect his favorite convictions from rational scrutiny by appeals to intution? What is to prevent someone from simply *asserting* the continuum hypothesis as an axiom of set theory, on the grounds that she and Cantor at least find it perfectly 'intuitive', and those who do not simply don't understand the concept of set? Is there any principled way of settling what can, and what cannot be said to be 'intuitive', in this non-psychological sense? The threat of scientifically obstructive dogmatism is of course the feature of Frege's position on geometry that so disturbed Hilbert. Now as I understand him, Gödel is arguing that there is a principled way of drawing this distinction. The concept of set is grounded in our intuitive grasp of the basic mathematical structure, that of a well-founded linear order. Extensions of the set concept at the outer limits of set theory are to be assessed with respect to their consequences for our understanding of this basic structure, and Gödel thinks that this provides constraint enough to ensure that the concept of set is scientifically manageable.



The idea is this. One ascends in parallel on each side of the diagram, with the demands of physical theory drawing one through naive physics, then a more developed physics including the first hints of elementary dynamics, then naive dynamics, then Newtonian physics etc.. This progression induces on the mathematical side the demands met by elementary arithmetic, then by full arithmetic (including the properties of the rationals), then elementary analysis and calculus, etc. etc. At each stage, the explanatory demands on theory are met in part by the invocation of extra objects, both at the mathematical and the physical level. Ultimately, one will be drawn into the transfinite by the successive series of demands imposed on mathematical theory.

It seems to me, though, that the discussion in Chapter One shows that there is a quite fundamental problem with this Gödelian picture. On the scientific side, one might well want to agree that one is drawn deeper and deeper into an ontology of unobservables simply in order to get a cognitively manageable grasp of the observable in all its detailed and evolving complexity. At each stage, the answer to the question, But why should I believe that there are φ's? is provided with the response, Because there are these observational facts (look) that are best explained by this theory which quantifies over φ's. Let us agree that this is a good answer - that, on the physical side, theory really does suck us into the extra ontology in this way. But it is apparent that the parallel question on the other side must, at some stage, get a very different answer. Once the bounds of the calculable in principle are reached, there is no longer, even in principle, the possibility of responding to the ontological question by pointing out the independently available computational data requiring explanation, for there are no such independently available computational data. At this boundary point, then, some additional argument is needed.

I cannot see that Gödel gives any hint of a plausible additional argument. The only suggestions I can find are those implicit in what he says concerning the meaningfulness of questions in transfinite set theory, the possibilities of new axioms rich in mathematical consequences etc. However, I cannot agree that the only way to account for the meaningfulness of those questions, or the implications of such axioms, demands the extension of a realist ontology beyond the Hilbertian bound, for it seems to me that the interpretability of transfinite set theories in the weak sense introduced in **Chapter One** is perfectly sufficient to provide what is needed - or at least, it seems to me that, if this is not so, then we are in need of a philosophical argument that neither Gödel, nor such neo-Gödeleans as Maddy, have provided, in order to explain to us *why* it is not so. The Quinean argument for scientific realism does not appear to me to offer anything apt to plug this gap, and the more obvious ways in which one might hope to plug it - for example, by appealing to richer doctrines in semantic theory - seem unlikely to be compatible with the metaphysical underpinnings of the Quinean argument for moderate mathematical realism.

There is, however, one part of the long passage from Gödel quoted above that suggests, if not an argument to plug this gap, then at least some suggestions as to the availability of mathematical data apt to motivate such an argument. For towards the end of that quotation, Gödel mentions his famous incompleteness results as providing grounds for the belief that the minimal demands of mathematical theory, even with respect to the most elementary parts of mathematics, will require ever more far reaching excursions into the Cantorian transfinite. This is an intriguing suggestion. It is high time that the Hilbertian position I have been outlining was confronted with the fact of incompleteness. That is the topic of **Chapter Three**.

Summary. Before proceeding to discuss the incompleteness phenomena, let me close this part of my thesis with a summary of the discussion so far.

I described the Hilbertian project as the clarification and justification of the mathematician's use of the actual infinite. I showed that this project was to be accomplished by a demonstration, at the center of which lies the Master Argument, that the actual infinite of classical mathematics, and in particular classical analysis, was never needed to prove finitary results. I showed that there is good reason to believe that this project can be carried out with respect to classical analysis, although it cannot be carried out for those parts of mathematics that depend upon the full transfinite theory of the cardinal and ordinal numbers.

I then attempted to provide some motivation for Hilbert's concern to clarify and justify the use of the actual infinite by giving some historical background to Hilbert's Programme. In so doing, I stressed the continuity of his concerns from the early work in geometry through the later consistency programme. I argued that his fundamental motivation was provided by a desire to protect the accepted practice of classical mathematics from the kind of proto-intuitionist attacks of Kronecker (and later Brouwer), as well as the radical realism of Frege. I showed that this led him to attempt to draw philosophical discussion of mathematics away from considerations grounded in general metaphysical theses developed independently of the actual practice of mathematics, towards considerations such as consistency, independence etc. which could be addressed by properly mathematical means. In so doing, I tried to reinforce the point that a proof of consistency, for Hilbert, served not so much to deflect some genuine worry that classical mathematics might be inconsistent, but rather to show that worries about our inability to deal with infinitistic notions were ungrounded.

However, I then argued that there is much confusion in Hilbert's attempts to accomplish these tasks. I said that the philosophical obligations he attempts to discharge greatly exceed anything strictly required for his purposes. I distinguished two notions of interpretation for mathematical theories, and argued that Hilbert's central purposes in the philosophy of mathematics can be accomplished by maintaining that the meaningfulness of mathematical theories is sufficiently grounded in interpretation in what I called the *thin* , purely internal sense. I said that the principle (**H1**) can be made to seem quite plausible if the acceptability of mathematical theories is regarded as answerable to interpretation in the thin sense.

But I also said that Hilbert's attempts to ensure the interpretability of mathematical theories in the thick sense have interesting features, and that a plausible philosophy of mathematics can incorporate those attempts. Deferring to the Quinean indispensability arguments, I claimed that some part of classical mathematics has a thick interpretation. Departing from Quine, I identified that part with the mathematics that is formalizable in conservative extensions of **PRA**. I attempted to do some justice to Hilbert's claims about a part of mathematics necessary for all 'scientific' thought, by arguing that PRA is a presupposition of representational theories of cognition, and thus basic to any attempt to theorize the notion of a mind independent reality. I then suggested that an analogue of the Quinean argument can be used to extend this moderate realism over all of that part of mathematics conservative over **PRA**, where the infinite may indeed be regarded as simply a unifying, simplifying

device, just as Hilbert claimed. I concluded Chapter One by attempting to argue against some richer notions of mathematical realism, as well as the instrumentalist construal of Hilbert's Programme.

In Chapter Two, I devoted some energy to arguing against the claims that finitary mathematics is in some way ontologically special - committed only to expressions, for example - or epistemologically special. I then attempted to explain further the special status allocated to finitary mathematics in Chapter One, by a more detailed account of the centrality of computation to any theory of cognition which drew upon Gödel's well-known discussion of mathematical intuition. But I denied that this discussion has any power to introduce entities into our ontology beyond those countenanced by the moderate realism accepted in Chapter One.

CHAPTER THREE:
The Incompleteness Theorems and Hilbert's Programme

Introduction: In this chapter, I discuss the implication for Hilbert's Programme of the two incompleteness theorems of Gödel.

The principal implication, of course, is this: Hilbert's Programme cannot be carried out with full generality. This is the conventional wisdom, and it seems to me to be correct. However, there has been surprisingly little consensus over just *how* the incompleteness theorems show that Hilbert's Programme cannot be carried out. Most often, it is the Second Incompleteness Theorem that is said to have struck the killing blow. But as we shall see, the Second Incompleteness Theorem is a very curious result, and its impact on the Hilbert Programme is less clear cut than one might think. Mindful of this, some of the more insightful of recent commentators have taken the view that it is in fact the First Incompleteness Theorem that poses the deepest problem for Hilbert.

This is my own view, and I shall give reasons in support of it in Section Four below. However, Michael Detlefsen, in his book on Hilbert's Programme and in subsequent work, has denied that *either* of the incompleteness theorems have any tendency to show that Hilbert's Programme cannot be carried out. Detlefsen's arguments will be discussed at some length.

Whilst I agree that the incompleteness theorems show that Hilbert's Programme cannot be carried out with full generality, however, I do not accept the view that this shows that the underlying Hilbertian philosophy of mathematics is in any way defective. Indeed, it seems to me that that underlying philosophy, properly understood, is greatly strengthened by the incompleteness phenomena. This, of course, was not Gödel's own view, for he took his discoveries to provide support for some stronger form of mathematical realism. I shall explain below just how implausible that Gödelian contention is.

Section One: The Diagonal Lemma and the First Incompleteness Theorem. It is not my intention to reproduce any of the standard proofs of incompleteness. Many detailed proofs reasonably accessible to non-mathematicians are readily available in the

literature.[152] However, it will be necessary to draw attention to some particular features of the standard proofs.

The heart of the logical facts uncovered by any proof of the First Incompleteness Theorem is the *diagonal lemma*. Here it is, stated with full generality:

> **(DIAG)** Let T be any extension of Q. Then for any formula $\varphi(x, v_1, \ldots, v_n)$ of L(T), there exists a formula $\psi(v_1, \ldots, v_n)$ of L(T) such that
>
> $$T \vdash (\forall v_1) \ldots (\forall v_n)(\psi(v_1, \ldots, v_n) \leftrightarrow \varphi(\ulcorner \psi \urcorner, v_1, \ldots, v_n))$$

This theorem is not proved in Gödel [1931]. Rather, Gödel (in effect) proves a special case of it, constructed with respect to a Principia Mathematica-type formalism. In the above generalized form, the diagonal lemma was first proved by Montague.[153] There is a detailed proof of a slightly less general form of the lemma in Boolos and Jeffrey [1990], and a succinct proof of the lemma as stated above in McGee [1991].[154]

As a special case of (DIAG), we have

> **(DIAG*)** With T as above, for any formula F(x) of L(T) with x alone free, there is a sentence G of L(T) such that
>
> $$Q \vdash G \leftrightarrow F(\ulcorner G \urcorner)$$

As a brief indication of the import of (DIAG*), suppose we introduce a predicate T(x) into the language of some consistent extension of Q^+ of Q, satisfying $T(\ulcorner S \urcorner) \leftrightarrow S$ for every sentence S (a *truth predicate* for Q^+). Using (DIAG*), we can find a sentence G such that $Q^+ \vdash G \leftrightarrow \neg T(\ulcorner G \urcorner)$, whence we will have $Q^+ \vdash T(\ulcorner G \urcorner) \leftrightarrow \neg T(\ulcorner G \urcorner)$, i.e. Q^+ is inconsistent. This is a formalized version of the Liar Paradox.

---

[152] Gödel [1931] is perhaps the clearest account of the specific result proved in that paper. For the generalized version of the First Incompleteness Theorem discussed here, however, Boolos and Jeffrey [1989] chapters 14, 15, and 28 is the best treatment known to me. Monk [1976] provides a very comprehensive discussion, but one intended for mathematicians. For the Second Incompleteness Theorem, Monk is excellent, as is Boolos [1979].

[153] See Montague [1962]. Few textbooks prove the Diagonal Lemma with full generality, and some standard treatments of the incompleteness theorems (for example, Kleene [1952] and Schoenfield [1967] do not mention it at all. Boolos [1979] gives a comprehensive treatment. For some details of the curious history of the Diagonal Lemma, see Smorynski [1981].

[154] See Boolos and Jeffrey [1990] pp170-180, and McGee [1991] 24-25.

Suppose now that T is some consistent, axiomatizable extension of Q. Then the relations xMy and xNy defined by

> (a) xMy iff x is the code number of a proof in T of the formula with the code number y
>
> (b) xNy iff x is the code number of a proof in T of the negation of the formula with the code number y

are recursive. Since these relations are recursive, there exist predicates μ and ν of L(T) which define M and N respectively in T.

We now let F(x) be the formula

> (c) $(\forall y)(\mu(y, x) \rightarrow (\exists z)((z < x \ \wedge \ \nu(z, x)))$.

By (DIAG), there exists a formula G such that

> $T \vdash G \leftrightarrow F(\ulcorner G \urcorner)$.

It can then be shown that

> (ROSS) Neither $T \vdash G$ nor $T \vdash \neg G$.

The sentence G constructed in this way is *undecidable* in T - for the proof, which is straightforward, see Boolos and Jeffrey [1991] pp285-287. This is a version of the First Incompleteness Theorem, due to Rosser.[155]

A theory T is said to be *ω-inconsistent* if, for some formula F(x), $T \vdash \neg F(k)$ for every natural number k, and $T \vdash (\exists x)F(x)$. T is ω-consistent if T is not ω-inconsistent. On the assumption that T is an ω-consistent extension of Q, a simpler construction will result in an undecidable sentence of L(T). This time, we let F(x) be the formula

> (d) $\neg(\exists y)\mu(y, x)$.

Then by (DIAG) there exists a sentence G of L(T) with

---

[155] See Rosser [1939].

(GÖD) $T \vdash G \leftrightarrow \neg(\exists y)\mu(y, \ulcorner G \urcorner)$.

Once again, the sentence produced in this way is undecidable in T - for details, see Boolos and Jeffrey [1990] pp287-288. This form of the First Incompleteness Theorem is due to Gödel.[156]

Neither (ROSS) nor (GÖD) is the First Incompleteness Theorem as I shall understand it in this essay, however. For our purposes, the First Incompleteness Theorem is what is established by (ROSS), (GÖD) and the detailed work on the arithmetization of syntax that makes (DIAG) provable; viz:

(Gödel 1) There is no consistent, complete, axiomatizable extension of Q.

The argument is simple. If T is any axiomatizable extension of Q, then T is capable of representing the recursive functions, whence (DIAG) is provable in T. If T is consistent, then we can effectively find a formula G as in (Ross) above (for example) for which neither $T \vdash G$ nor $T \vdash \neg G$, whence T is incomplete. The important point to notice here, though, is that the conditions mentioned in the formulation of (DIAG) and (Gödel 1) - that T be a consistent extension of Q, that F(x) be a formula with x alone free etc. - are all readily and naturally translatable into mathematical idiom, in a way which permits the ready generalization of the particular result proved in Gödel [1931] to a wide range of mathematical theories. It is therefore reasonably obvious that there is a genuinely mathematical fact, of which (Gödel 1) is a natural transcription into ordinary English. We shall see that the situation with respect to the Second Incompleteness Theorem is rather different.

One final point. All the results mentioned in this section are purely syntactic in character. No claim is made here about the content of the underivable sentences constructed via (DIAG), and no such claim is made by Gödel about the sentence '17 Gen r' which he shows to be underivable in a Principia-style formal system in Gödel [1931]. I do not say that such sentences have no content: I am simply observing that the logical facts surrounding the First Incompleteness Theorem can be stated without appeal to their content.

---

[156] See Gödel [1931].

Section Two: Provability Predicates and the Second Incompleteness Theorem. With T still any consistent extension of Q, the reader will recall that the relation $P_T(x, y)$ that holds of numbers n, m just in case n is the code of a proof in T of the formula with code number m is computable, therefore recursive, therefore representable in Q. We shall now need to be a little more precise about the notion of (formal) proof. A proof in T, we shall say, is a finite sequence of formulas of L(T), separated by commas (in the encoding of syntactic objects, the comma will have been assigned a code number), each one of which is either an axiom of T or some formula B where A and $A \rightarrow B$ are earlier formulas in the sequence. *This is the standard notion of proof*, and one of the objectives of arithmetization, indeed of proof theory as a whole, is to investigate its properties.

Under any reasonable encoding of the syntax of T, each proof of T will therefore have a unique code. By formalizing the definition just given in the natural way, and by speaking of codes rather than expressions, and using any standard device for encoding finite sequences, we can define a *particular formula* Pr(x, y) (henceforth suppressing the relativization to T) which represents, or, as we shall now say, *numeralwise expresses* the relation 'x is the code of a proof in T of the formula with the code y'.

With Pr(x, y) as just defined, we now define Bew(y) to be the formula $(\exists x)Pr(x, y)$.

Suppose now that we have $T \vdash A$ for some formula A. Then there is a proof in T of A, to which is assigned some code n. Then $T \vdash Pr(\mathbf{n}, \ulcorner A \urcorner)$, so $T \vdash (\exists x)Pr(x, \ulcorner A \urcorner)$, i.e. $T \vdash Bew(\ulcorner A \urcorner)$. Summarizing:

(DER 1) If $T \vdash A$, then $T \vdash Bew(\ulcorner A \urcorner)$.

Suppose now that we have $T \vdash A$ and $T \vdash (A \rightarrow B)$. Then by writing down the proof of $(A \rightarrow B)$, followed by a comma, followed by a proof of A, followed by a comma, followed by B, we obtain a sequence that is a proof of B. This reasoning can be formalized in any appropriate T, and it then yields

(DER 2) $T \vdash Bew(\ulcorner A \rightarrow B \urcorner) \rightarrow (Bew(\ulcorner A \urcorner) \rightarrow Bew(\ulcorner B \urcorner))$.

A third feature of $Bew(\ulcorner x \urcorner)$, which plays a very important role in the proof of the Second Incompleteness Theorem, is this:

(DER 3) $T \vdash Bew(\ulcorner A \urcorner) \rightarrow Bew(\ulcorner Bew(\ulcorner A \urcorner) \urcorner)$.

(Notice the resemblance between **(DER 3)** and the characteristic axiom of the modal system **K4**, and also between **(DER 3)** and the controversial K-K principle in epistemology.[157]) This time the verification that $Bew(\ulcorner x \urcorner)$ satisfies this condition is lengthy.

(DER 1) - **(DER 3)** together make up the Löb *Derivability Conditions*. They are simplifications of the Hilbert-Bernays derivability conditions first established for extensions of **Q** in Hilbert and Bernays [1939], and they play the central role in any generalized version of the Second Incompleteness Theorem.[158] The proof that the particular predicate which numeralwise expresses derivability whose construction was sketched above satisfies **(DER 1)** - **(DER 3)** is complex, and several versions are available in the literature.[159] I do not intend to reproduce the details.

However, I do wish to comment on one matter which an examination of the details makes clear. It is highly misleading to speak of the Derivability Conditions - as Detlefsen does, for example - as *constraints* that a formula that represents the derivability relation must meet, if a generalized version of the Second Incompleteness Theorem is to be obtained. For this suggests that **(DER 1)** - **(DER 3)** are being imposed upon the notion of formal derivability in the interests of getting a generalized Second Incompleteness Theorem, and that gets matters the wrong way round. The point is that **(DER 1)** - **(DER 3)** *are in fact satisfied* by the formula which represents derivability in any appropriate system, and *which is constructed in the way which mimics our natural understanding of derivability*. Thus, a *derivation* is a *finite sequence* of *formulas*, each of which is either an *axiom* or is *B where A and A* $\rightarrow$ *B are antecedent formulas*, or . . . etc. There is, for each of these italicized expressions, a natural formalization, and a good arithmetization of these natural formalizations will mimic the resulting theorems of syntax as theorems of arithmetic. The

---

[157] The K-K principle says that, if S knows that P, then S knows that S knows that P.

[158] Strictly and literally, this is false. In Jeroslow [1973] it is demonstrated that a general version of the Second Incompleteness Theorem can be proved without reliance on **(DER 2)**. However, since the main controversy over the Derivability Conditions concerns **(DER 1)** and, especially, **(DER 3)**, I neglect this complication.

[159] The original proof is in Hilbert and Bernays [1939]. There are detailed, up to date treatments in Monk [1976] Chapter 17, and Boolos [1979] Chapter 2. Many more details are given in the forthcoming second edition of Boolos [1979]. Finally, there is a vivid, if impressionistic discussion in Chapter 0 of Smorynski [1985].

result of all this is an expression which represents derivability, and which can then be shown to satisfy **(DER 1)** - **(DER 3)**. No 'constraints' have to be 'imposed': rather, those conditions required for a generalized proof of the Second Incompleteness Theorem are in fact satisfied by the natural representation of the metamathematical notions involved. The import of this will, I hope, become a little clearer shortly.

A formula F(x) (with x alone free) which numeralwise expresses provability, and which satisfies **(DER 1)** - **(DER 3)** for any sentences A, B of the language of **T** is called a *provability predicate* of **T**.

We can now state the central theorem of this section:

> **(Löb)** If F(x) is a provability predicate for **T**, then for any sentence A, if $\mathbf{T} \vdash F(\ulcorner A \urcorner) \rightarrow A$, then $\mathbf{T} \vdash A$.

For the proof of **(Löb)**, see Löb [1955], or Boolos and Jeffrey [1989] pp187-188. It is instructive to consider the formula that stands to **(Löb)** in the same relation as **(DER 3)** stands to **(DER 1)**, viz.

> **(FLöb)** $\mathbf{T} \vdash F(\ulcorner F(\ulcorner A \urcorner) \rightarrow A \urcorner) \rightarrow F(\ulcorner A \urcorner)$.

In the modal system **G** which has the modal analogues of **(DER 2)** and **(FLöb)** along with all tautologies as axioms, and has modus ponens and necessitation as rules of inference, the modal analogue of **(DER 3)** is *derivable*. **G** is the system of *provability logic*, and is of particular interest, as modal systems go, in virtue of the following fact. Let φ be any function assigning sentences of PA to sentence letters of **G**. For each sentence F of **G**, define F^{φ} as follows:

> $p^{\varphi} = \varphi(p)$ for p a sentence letter.
> $\bot^{\varphi} = \mathbf{0} = \mathbf{1}$
> $(A \rightarrow B)^{\varphi} = (A^{\varphi} \rightarrow B^{\varphi})$
> $(\Box A)^{\varphi} = \mathrm{Bew}(\ulcorner A \urcorner)$

Then the following can be shown

(Arithmetical Soundness of G):

if $G\vdash A$, then for all φ, $PA\vdash A^{\varphi}$.[160]

With (Löb) in hand, the proof of the Second Incompleteness Theorem is agreeably simple:

> **(Gödel 2)** If F(x) is a provability predicate for T, and if T is consistent, then not: $T\vdash \neg F(\ulcorner 0=1 \urcorner)$.

Proof: Suppose F(x) is a provability predicate for T, and suppose $T\vdash \neg F(\ulcorner 0=1 \urcorner)$. Then by the sentential calculus, $T\vdash F(\ulcorner 0=1 \urcorner) \rightarrow 0 = 1$, whence by (Löb) $T\vdash 0 = 1$, and T is inconsistent.

Let us now look more carefully at relations Pr(x, y) that numeralwise express the provability relation of a theory. Recall that an n-place relation $R(x_1, \ldots, x_n)$ amongst natural numbers is said to be *numeralwise expressible* in the system T iff there is a wff $F(x_1, \ldots, x_n)$ of L(T) (with n free variables) such that, for any natural numbers $k_1, \ldots, k_n$,

> (a) if $R(k_1, \ldots, k_n)$ holds, then $T\vdash F(\mathbf{k_1}, \ldots, \mathbf{k_n})$
> (b) if $R(k_1, \ldots, k_n)$ does not hold, then $T\vdash \neg F(\mathbf{k_1}, \ldots, \mathbf{k_n})$.

For example, the number theoretic relation 'less than' is numeralwise expressible in PA by the formula $x < y$. For, if $k_1 < k_2$, then $k_2 = k_1 + k_3$, where $k_3 \neq 0$. It is easy to see that we have $PA\vdash \mathbf{k_2} = \mathbf{k_1} + \mathbf{k_3}$ and $PA\vdash \mathbf{k_3} \neq \mathbf{0}$, whence $PA\vdash (\exists x)(x \neq 0 \wedge x + k_1 = k_2)$ - i.e. we have $PA\vdash \mathbf{k_1} < \mathbf{k_2}$. On the other hand, if $k_2 \leq k_1$, it is easy to see that we have $PA\vdash \mathbf{k_2} \leq \mathbf{k_1}$, i.e. $PA\vdash \neg \mathbf{k_1} < \mathbf{k_2}$.

This example will make it clear that the demand that a formula numeralwise express some number theoretic relation imposes only a very weak constraint on formulas. If a relation is numeralwise expressible in a theory T at all, it is numeralwise expressible in that theory by infinitely many different but co-extensive formulas. There are therefore infinitely many formulas of Q that numeralwise express the relation P(x, y) that holds of numbers x, y just in case x is the code of a proof in Q of the formula with code number y.

---

[160] See Boolos and Jeffrey [1989] Chapter 27 for the proof. For more details, see Boolos [1979].

This is important for the following reason. The particular formula Pr(x, y) whose construction was sketched above yields a formula Bew(y) which in fact satisfies the derivability conditions, and is therefore a provability predicate. But there are a great many alternative formulas Pr*(x, y) which also numeralwise represent the provability relation yet for which Bew*(y) = (∃x)Pr*(x, y) is *not* a provability predicate. Here is an example. With Pr(x, y) as above, define Dr(x, y) to be the formula (Pr(x, y) ∧ y ≠ ⌜0 = 1⌝).[161] Dr(x, y) numeralwise expresses provability, since Dr and Pr are coextensive. But Dew(y) = (∃x)Dr(x, y) is not a provability predicate, since it violates **(DER 2)**. .

Interestingly, the formula (c) above that features in Rosser's version of the First Incompleteness Theorem yields another example of this same phenomenon. With μ(x, y) a formula that numeralwise expresses the provability relation, the formula

$$(c^*)\quad \mu(x, y) \wedge \neg(\exists z)(z < y \wedge \mu(z, \dot{-}(y)))$$

says something like 'x is a proof of y, and there is no shorter proof of the negation of y'. In a consistent theory, (c*) must numeralwise express what μ(x, y) does. However, the predicate Bew*(y) constructed from (c*) as Bew(y) was constructed from Pr(x, y) is also not a provability predicate. Not only that, however, the sentence ¬Bew*(⌜0 = 1⌝), with Bew*(y) constructed from (c*), is in fact *derivable* in any consistent extension of Q.

This is rather a curious situation, then. We have sentences S and S*, similarly constructed from co-extensive predicates, each with at least some sort of a claim to express consistency, such that T⊢ S* but not T⊢ S. Auerbach [1985] provides an even simpler, and illuminating example of the same phenomenon.[162] Define μ*(x, y) as μ(x, y) ∧ ¬μ(x, ⌜0 = 1⌝). In a consistent system, μ and μ* numeralwise express the same relation. But whereas the instance of the consistency schema

$$\neg(\exists x)\varphi(x, \ulcorner 0 = 1 \urcorner)$$

constructed with μ, viz.

$$\neg(\exists x)\mu(x, \ulcorner 0 = 1 \urcorner)$$

---

161 'Dr' abbreviates 'droof' which abbreviates 'Dreben proof'.

162 See Auerbach [1985] p343.

is not derivable, the instance constructed with μ^*, viz.

$$\neg(\exists x)[\mu(x, \ulcorner 0 = 1 \urcorner) \wedge \neg\mu(x, \ulcorner 0 = 1 \urcorner)$$

is derivable, since it is a theorem of logic.

Still more peculiar phenomena can be wrung out of (Gödel 2) by exploiting the properties of co-extensive formulas which numeralwise represent the same relation. In the passage in which provability predicates were introduced above, for example, we spoke of transcribing in L(T) the definition of 'proof' 'in the natural way'. In a subject noted for rigor and robustness of argumentation, it may seem odd to have to resort to stipulations as to what is and what is not 'natural' in the formalization of some notion; but in this case the oddity is essential. For with a little ingenuity, it is possible to define a formula $\varphi(x)$ which numeralwise expresses the property of being an axiom of T, and therefore there are infinitely many co-extensive formulas which have this same capacity. Now, our predicate Pr(x, y) from which we constructed a provability predicate Bew(y) was the result, as we have said, of transcribing the notion of proof, and therefore the notion of axiom 'in the natural way'. However, we have just seen that there are infinitely many ways of formally expressing the notion 'proof of T', and therefore infinitely many Pr*(x, y) which differ only with respect to the formula chosen to numeralwise represent the property of being an axiom of T. All these different ways of numeralwise representing the axioms of T are of course ways of presenting exactly the same formal system T, in the ordinary sense in which T is identified with its theorem set. But now we can define infinitely many different predicates which numeralwise express the relation P(x, y), varying only with respect to how the axioms of T are presented by the formula which numeralwise expresses the property of being an axiom of T.

Feferman, in his important study of the Second Incompleteness Theorem, shows how to construct deviant 'consistency' sentences in a uniform way for a large class of systems with respect to which a version of (Gödel 2) is provable.[163] Feferman's deviant proof predicates are formalized as the transcriptions of a carefully chosen, and somewhat peculiar *presentation* of the axioms of the system in question, and his consistency sentences are

---

[163] See Feferman [1960]. This seems as good a place as any to mention that the work of Feferman and Jeroslow (mentioned below) owes much to the stimulation of remarks found scattered throughout the work of Kreisel - see e.g. the influential article Kreisel [1965].

then natural transcriptions of the standard notion of consistency as $\neg(\exists x)Pr^*(x,\ulcorner 0=1\urcorner)$. But, as one might expect from the discussion above, some of these ways of 'presenting' the axioms of T are peculiar enough to ensure the provability of a 'consistency' sentence for T. By this I mean that there are formulas Pr*(x, y) which can be used to construct a 'consistency' sentence $\neg(\exists x)Pr^*(x,\ulcorner 0=1\urcorner)$ which *is* provable in consistent extensions of Q. The consequence is once again that, for a given consistent extension T of Q, there will be formulas S and S*, constructed in the same way from co-extensive predicates, each of which having some claim to express the consistency of T, differing only with respect to how the axioms of T have been 'presented' in their construction, such that $T\vdash S^*$ but not $T\vdash S$. One moral of this, then, is that some of the metamathematical results provable about a given system T will depend upon the precise conditions satisfied by the choice of a particular predicate P(x,y) to numeralwise express the proof relation.

It will be instructive to examine this work of Feferman in greater detail, for two reasons (not counting its intrinsic interest). Firstly, it is of some importance to the philosophical issues that will occupy us in the following section. Secondly, and more important, it will allow me to give a slightly more concrete account of what I speak of above as the natural way of formalizing and arithmetizing metamathematical notions.

The syntax of the systems Feferman deals with in Feferman [1960] is arithmetized in very much the normal way, with syntactic primitives coded by numerals, and syntactic operations (such as substitution for free variables) characterized by recursive number theoretic functions. Relative to a given non-logical vocabulary K, we have e.g. 'Const' used to denote the (recursive) set of non-logical constants of K, 'Fm_K' to denote the formulas of K, 'Tm_K' the terms of K, 'St_K' the sentences of K, 'Sq' an arbitrary sequence of formulas of K, etc. A formal system becomes a pair T = <A, K> where $A \subseteq St_K$ and $K \subseteq Const$. The system L_T = <0, K> is then the pure logic of T. For each n, A/n is the set of all $q \in A$ with $q \leq n$. T/n = <A/n, K> is then a *finite subsystem* of T = <A, K>.

The idea is then to arithmetize the relation of logical derivability by simply copying (directly, 'in the natural way') its ordinary definition in this arithmetized syntax. The result is this:

Prf_A is the binary relation such that: for any φ, ψ, $Prf_A[\varphi, \psi]$ iff $\psi \in Sq$ and $\varphi = (\psi)_{L(\psi)-1}$ and for each $i < L(\psi)$, $(\psi)_i \in Fm_K$ and either (a) $(\psi)_i \in Am_K$, or (b) $(\psi)_i \in A$, or (c) for some $j, k < i$, $(\psi)_k = (\psi)_j \rightarrow (\psi)_i$.

This says: for any formulas φ, ψ; ψ is a (representation of a) proof of φ from the sentences A iff ψ is a (representation of a - I shall henceforth elide this) sequence of formulas, φ is the end formula of the sequence, and each formula in the sequence is either an axiom, an element of A, or is C where B and $B \rightarrow C$ are antecedent formulas in ψ. It is the natural transcription of the notion of *derivation* into formalese.

So far, this is just as one would expect. Feferman is then able to develop familiar results about derivability, including the deduction theorem (Feferman's 2.2) and a (trivial) kind of 'compactness' theorem (2.3) asserting that anything provable from the axioms in A is provable from a finite subset of A.

The systems **Q** and **PA** are then introduced. Feferman also considers systems that are PR-extensions of **PA**, where a PR-extension of a theory **T** is the result of adding finitely many new function symbols to L(T), together with primitive recursive equations defining those symbols as new axioms. A predicate of **T** which is constructed without any use of unbounded quantifiers is then called a *PR-predicate* of **T**.

The necessary results on numeralwise expressibility are then demonstrated as theorems (3.4 (i)) and (3.4 (ii)). Feferman's term for 'numeralwise express' is 'binumerate', and he also uses 'numerate' to express what is otherwise expressed by 'weakly express' or 'weakly define'. Theorem (3.4) then tells us that, (i) for each n + 1-ary p.r. function φ there exists a PR-extension **PA'** of **PA** with a term t (in n free variables) such that the formula $t(v_1, \ldots, v_n) = v_{n+1}$ numerates φ in **PA'**, and every such formula numerates a p.r. function; and (ii) similarly for each n+1-place relation, with 'bi-numerates' for 'numerates'. The elimination techniques used in Gödel [1931] are then used to show that (bi)-numeration in **PA'** can always be replaced by (bi)-numeration in **PA**.

The notion of a PR-formula is then introduced, as a formula $f(x) = 0$ in some PR-extension of **PA**. Given a PR-predicate F of **T**, the result $(\exists x_1) \ldots (\exists x_n)F$ of prefixing F with a string of existential quantifiers is called an RE-formula of **T**. (Given the standard primitive recursive pairing function, such initial strings of existential quantifiers are reducible to a

single existential quantifier without loss of generality.) The classes of PR-formulas and RE-formulas are primitive recursive, and closed under conjunction, disjunction, bounded quantification, and (in the case of the PR-formulas) negation. These constructions, which are still pretty much as one would expect, culminate in Theorems 3.11 and 3.12, which establish in Feferman's terms the familiar fact that the p.r. functions are representable in **Q**, and thus in all consistent extensions of **Q**.

We have arrived at Section Four of Feferman's paper, which is the heart of the matter from our point of view. Feferman now introduces a particular p.r. extension **M** of **PA**, together with a new piece of notation. **M** is best understood by considering an example of the notation. Let Exp be the p.r. exponentiation function $\mathrm{Exp}(n, m) = n^m$. Corresponding to this function in **M** is a two-place function symbol e(x, y), and an axiom

$$(\forall x)(e(x, 0) = 1 \wedge (\forall y)(e(x, y') = e(x, y) \times x))$$

Then Feferman writes

$$\underline{n}^{\underline{m}} \quad \text{for} \quad e(n, m).$$

(In fact, Feferman uses a dot where I have used underlining. This is attributable to the local limitations of word processors, and I hope that it will cause no confusion.) Therefore, '$\underline{n}^{\underline{m}}$' denotes a term of **M** that represents exponentiation in **M**. With this notation, then, one can indicate arithmetizations of metatheorems by writing down the metatheorems in ordinary mathematical notation, prefixing '$\mathbf{M}\vdash$', and underlining the mathematical/logical operators and predicates involved.

More precisely, for each n-place p.r. function is associated an n-place function symbol f of **M**. Let 'F' be the ordinary mathematical notation for this function. Then '$\underline{F}$' is the metamathematical notation for the term $f(x_1, \ldots, x_n)$ of **M**, and

$$\underline{F}(t_1, \ldots, t_n) = f(t_1, \ldots, t_n).$$

We can extend this device to predicates via their characteristic functions in the usual way. Thus, '$\underline{\mathrm{Prime}}(y)$' denotes a predicate of **M** satisfied by primes only, whilst '$\underline{\mathrm{Sb}}(F[a/t])$' e.g. denotes a term of **M** for the result of subbing t for a in F, and '$\underline{\mathrm{Ax}}_{\mathbf{K}}$' denotes a term of **M** for the set of axioms of **K**.

What is not yet clear, however, is how to express in **M** the idea of *provability from an arbitrary set A of axioms*, for it is not clear how in general we are to express the idea of membership in **A** in **M**. Feferman writes:

> . . . if we look back at the definition of Prf_A, we see that the only concept which enters in it which we may not be able to express in **M** is that of membership in A. Even if we know that A is recursive, or primitive recursive, *we are still faced with the further problem of choosing one from among the many numerations of A in L.* [164]

Feferman's way of getting around this difficulty is the essence of his approach to the problem of getting a generalized version of the Second Incompleteness Theorem, and it is also what enables him to define a 'deviant' notion of consistency according to which PA (for example) can 'prove' its own (Feferman)-consistency. He continues:

> In order not to prejudice the investigations, we therefore take the following initially non-committal approach. Let φ be a formula with one free variable x. We shall define a formula Prf_φ with two free variables x, y which will express, when $\varphi(x)$ is read as expressing that x belongs to A, that y is a proof from A = <A, K> of x.[165]

This is the definition:

> **(Fef 4.1)** Let φ be a formula of **M**, and let u, v, w be the first three variables not free in φ and distinct from x, y, z. We take
>
> $\underline{Prf_\varphi}$
>
> to be the following formula of K_0
>
> $$[\underline{Sq}(y) \wedge \underline{L}(y) \neq 0 \wedge$$
> $$(\forall u)[u < \underline{L}(y) \rightarrow \underline{Fm_K}((y)_u) \wedge$$
> $$[\underline{Ax_K}((y)_u) \vee \varphi((y)_u) \vee (\exists v)(\exists w)(v < u \wedge w < u \wedge (y)_{\underline{v}} = (y)_{\underline{v}} \underline{\rightarrow} (y)_u)]] \wedge$$
> $$x = (y)_{\underline{L(y) \dot{-} 1}}]^{(M)}$$

---

[164] Feferman [1960] p58, my emphasis.
[165] Feferman [loc cit].

This is not easy to read, but it is in fact an arithmetized version of the original definition of Prf_A, with φ for A. It is the natural arithmetized formalization of the notion of provability.

Thus defined, $\underline{Prf}_\varphi$ numeralwise expresses Prf_A, but it also has two other features worth remarking upon. In the first place, Feferman's dot notation is effectively defined. More importantly though, $\underline{Prf}_\varphi$ depends upon φ, and φ is of course a formula *in one free variable*. This enables Feferman to prepare the way for the introduction of a deviant notion of consistency, via a deviant way of 'presenting' the arithmetic axioms. The crucial constructions proceed as follows:

> **(Fef 4.2)** Let φ be a formula of M, with $Fv(\varphi) = \{x\}$, and φ' the formula $\varphi(x) \wedge x \leq z$. We put
>
> $$\underline{Prf}_{\varphi/z}$$
>
> to be the formula
>
> $$\underline{Prf}_{\varphi'}.$$

$\underline{Prf}_{\varphi'}$ is therefore a kind of restriction on $\underline{Prf}_\varphi$ - in effect, a restriction to the axioms preceding φ in some ω-ordering of axioms. The Feferman proof predicate is then introduced in the following two definitions:

> **(Fef 4.2)** Let A be a finite set, $A = \{k_0, \ldots, k_{n-1}\}$. Suppose that $k_0 < \ldots < k_{n-1}$. By [A] we mean the formula $x \neq x$, in case $n = 0$, and the formula
>
> $$x = k_0 \vee \ldots \vee x = k_{n-1}$$
>
> if $n > 0$.

(Thus $\underline{Prf}_{[A]}(x, y)$ expresses the proof relation relativized to the *finite* set A, and $\underline{Prf}_{[0]}(x, y)$ expresses logical provability.)

(Fef 4.3) For any formula φ of M, $\underline{Pr}_{\varphi}$ and $\underline{Pr}_{\varphi/z}$ are the formulas of K_0 defined as follows:

(i) $\underline{Pr}_{\varphi}(x) = (\exists y)\underline{Prf}_{\varphi}(x, y)$
(ii) $\underline{Pr}_{\varphi/z}(x) = (\exists y)\ \underline{Prf}_{\varphi/z}(x, y)$.

The extensional correctness of these definitions is readily proved (see Theorem 4.4), and some elementary facts about provability can be established with φ occurring schematically, including

(Fef 4.6) Let φ be formula of M, $Fv(\varphi) \subseteq \{x, z\}$.

(i) $M \vdash \underline{Pr}_{\varphi}(x) \rightarrow \underline{Em}_K(x)$.
(ii) $M \vdash \underline{Ax}_K(x) \rightarrow \underline{Pr}_{\varphi}(x)$.
(iii) $M \vdash \varphi(x) \wedge \underline{Em}_K(x) \rightarrow \underline{Pr}_{\varphi}(x)$.
(iv) $M \vdash [\underline{Pr}_{\varphi}(x) \wedge \underline{Pr}_{\varphi}(x \dot{\rightarrow} y)] \rightarrow \underline{Pr}_{\varphi}(y)$.
(v) If ψ is any formula of M, $y \in Fv(\psi)$, then
$M \vdash (\forall x)[(\underline{Ax}_K(x) \rightarrow \psi(x)) \wedge (\varphi(x) \wedge \underline{Em}_K(x) \rightarrow \psi(x))] \wedge$
$(\forall x)(\forall y)[\underline{Em}_K(x) \wedge \underline{Em}_K(y) \wedge \psi(x) \wedge \psi(x \dot{\rightarrow} y) \rightarrow \psi(y)] \rightarrow$
$(\forall y)(\underline{Pr}_{\varphi}(x) \rightarrow \psi(x))$.

Feferman then shows that the *first two* of the Hilbert-Bernays derivability conditions follow from **(Fef 4.6)** and the extensional correctness of the definition of $\underline{Pr}_{\varphi}$, with φ occurring schematically. (The first of the Hilbert-Bernays derivability conditions is our **(DER 1)**, essentially, whilst the second says that, if $\neg\psi(v)$ is derivable with v free, then $\neg\psi(\mathbf{n})$ is derivable for each n.) The proofs of these facts depend essentially only upon the ability of M to follow an inductive definition, for the metatheorems are proved by mimicking in M, with the dot notation, constructive proofs of the original theorems.

Two other facts about $\underline{Pr}_{\varphi}$, important in the proof of the Second Incompleteness Theorem, as established as parts of

(Fef 4.7) Let α, β be formulas of M, $Fv(\alpha) \cup Fv(\beta) \subseteq \{x, z\}$. Then

(i) $M \vdash (\forall x)(\beta(x) \wedge \underline{Em}_K(x) \rightarrow \alpha(x)) \rightarrow (\forall x)(\underline{Pr}_{\beta}(x) \rightarrow \underline{Pr}_{\alpha}(x))$
(ii) $M \vdash \underline{Pr}_{\underline{Pr}_{\alpha}}(x) \leftrightarrow \underline{Pr}_{\alpha}(x))$

A notion of consistency is defined, once more in the natural way:

(Fef 4.9 (ii)) Let φ be a formula of M, $Fv(\varphi) \subseteq \{x, z\}$. Then

$$\underline{Con}_\varphi = (\forall x)[\underline{Fm}_K(x) \rightarrow \neg \underline{Pr}_\varphi(x) \vee \neg \underline{Pr}_\varphi(\ulcorner x \urcorner)]^{(M)}$$

Up to the definition of $\underline{Pf}_\varphi$, everything here has been as one would expect. But exactly which notion of proof it is that is picked out by the definition of $\underline{Pf}_\varphi$ depends upon the interpretation of the parameter φ. Consequently, exactly *which* notion of consistency it is that is being defined in (Fef 4.9 (ii)) - although surely *a* notion of consistency is being defined - is not yet clear, as we shall soon see.

Here are three theorems about $\underline{Con}_\varphi$

(Fef 4.10) Suppose that φ, ψ are formulas of M, $Fv(\varphi) = Fv(\psi) = \{x\}$.

(i) For any particular $\beta \in Fm_K$,
$$M \vdash \underline{Con}_\varphi \leftrightarrow \neg \underline{Pr}_\varphi(\beta \wedge \neg \beta)$$
(ii) $M \vdash \underline{Con}_\varphi \leftrightarrow \neg(\forall z)\underline{Con}_{f/z}$
(iii) $M \vdash (\forall x)(\beta(x) \wedge \underline{Fm}_K(x) \rightarrow \varphi(x)) \rightarrow (\underline{Con}_\varphi \rightarrow \underline{Con}_\beta)$

Generalized versions of the incompleteness theorems now follow (in Feferman's section 5). First, we have a version of the restricted Diagonal Lemma (DIAG*) above:

(Fef 5.1) Lemma: let $\psi \in Fm_{K0}$, $Fv(\psi) \subseteq \{x\}$. Then we can effectively find $\varphi \in Fm_{K0}$ such that

$$Q \vdash \varphi \leftrightarrow \psi(\ulcorner \varphi \urcorner)$$

The proof is essentially that given in Boolos and Jeffrey [1989] p173, transcribed into the underlining notation of M. With the aid of a definition

(Fef 5.2) For each $\varphi \in Fm_{K0}$, with $Fv(\varphi) \subseteq \{x\}$, we take μ_φ to be the sentence associated with $\psi = \neg \underline{Pr}_\varphi$ in **(Fef 5.1)**, such that

$$Q \vdash \mu_\varphi \leftrightarrow \neg \underline{Pr}_\varphi (\mu_\varphi)$$

this yields a version of **(Gödel 1)**, with

$$\text{not } T \vdash \mu_\varphi$$

for any φ numerating any consistent extension of **Q**, and

$$\text{not } T \vdash \neg \mu_\varphi$$

for any φ numerating any ω-consistent extensions of **Q** as well (by Rosser's construction). Notice that this can be stated without restrictions on φ.

For the Second Incompleteness Theorem, though, restrictions are required upon the formula φ enumerating the axioms of the system, and these restrictions in effect play a role akin to that of the crucial third derivability condition **(DER 3)** in Feferman's approach. To state the required restrictions, Feferman first proves as (Theorem 5.4) a metamathematized version of this theorem:

> If ψ is any *bounded prenex formula* (BPF), and $Fv(\psi) \subseteq \{v1, \ldots, vn\}$, then for any $k_1, \ldots, k_n$
>
> if $\psi(k_1, \ldots, k_n)$, then $Q \vdash \psi(\mathbf{k}_1, \ldots, \mathbf{k}_n)$

(proved earlier as theorem 3.10), which assures us that any true BPF theorem is provable in **Q**. Theorem **(Fef 5.4)** therefore tells us that this fact about **Q** *is itself provable in M*. As a corollary of this,

(Fef 5.5) Suppose that $\varphi \in St_{K0}$ is such that for some $\psi \in BPF$

$$Q \vdash \varphi \leftrightarrow \psi$$

Then

$$PA \vdash \varphi \rightarrow \underline{Pr_{\varphi}}(\varphi)$$

(because we have $M \vdash \underline{Pr_{\varphi}}(\varphi \leftrightarrow \psi)$ by earlier results, which together with the formalized version of the above fact about Q yields $PA \vdash \underline{Pr_{\varphi}}(\varphi) \leftrightarrow \underline{Pr_{\varphi}}(\psi)$, whence $PA \vdash \varphi \rightarrow \underline{Pr_{\varphi}}(\varphi)$ if $Q \vdash \varphi \leftrightarrow \psi$.) This yields a version of the Second Incompleteness Theorem, in the form

> **(Fef 5.6)** Suppose that $A = \langle A, K \rangle$ is a consistent axiom system with $PA \subseteq A$. Suppose that φ is an *RE-formula* which numerates A in S, where $Q \subseteq S \subseteq A$. Then
>
> $$A \vdash \underline{Con_{\varphi}} \leftrightarrow \mu_{\varphi}$$
>
> and hence
>
> $$\text{not } A \vdash \underline{Con_{\varphi}}$$

The proof is worth pausing over, for it makes clear that it is the restriction on φ, together with theorems **(Fef 5.4)** and **(Fef 5.5)**, which here play the role played by the crucial third derivability condition **(DER 3)** in the standard generalized proof of the Second Incompleteness Theorem. First, we have

$$\textbf{(1)}\ Q \vdash \neg\mu_{\varphi} \leftrightarrow \underline{Pr_{\varphi}}(\mu_{\varphi})$$

by the construction of μ_{φ}. Since every RE-formula is equivalent to some BP-formula, we can appeal to **(Fef 5.5)** and derive

$$\textbf{(2)}\ PA \vdash \neg\mu_{\varphi} \rightarrow \underline{Pr_{\varphi}}(\ulcorner\neg\mu_{\varphi}\urcorner).$$

Q is finite, and each $\beta \in Q$ is provable in A by hypothesis, so

(3) $S \vdash [Q](x) \rightarrow \underline{Pr}_\varphi(x)$

Then from (Fef 4.7) (i) and (ii)) above, we derive

(4) $A \vdash \neg\mu_\varphi \rightarrow \underline{Pr}_\varphi(\neg\mu_\varphi)$

whence

(5) $A \vdash \underline{Con}_\varphi \wedge \neg\mu_\varphi \rightarrow \underline{Pr}_\varphi(\neg\mu_\varphi)$

which is

(6) $A \vdash \underline{Con}_\varphi \wedge \neg\mu_\varphi \rightarrow \mu_\varphi$

by construction, and therefore

(7) $A \vdash \underline{Con}_\varphi \rightarrow \mu_\varphi$.

Conversely, since

(8) $PA \vdash \neg\underline{Pr}_\varphi(\mu_\varphi) \rightarrow \underline{Con}_\varphi$

we have

(9) $PA \vdash \mu_\varphi \rightarrow \underline{Con}_\varphi$

and so

(10) $A \vdash \underline{Con}_\varphi \leftrightarrow \mu_\varphi$

Suppose now

(11) $A \vdash \underline{Con}_\varphi$.

Then by (10),

(12) $A \vdash \mu_\varphi$

contra the First Incompleteness Theorem (Fef 5.3). Therefore,

(13) Not: $A \vdash \underline{Con}_\varphi$.

Feferman then writes:

> The main feature of [(Fef 5.6)] which we wish to bring attention to is that, in contrast to [(Fef 5.3)] *it is not stated for arbitrary numerations φ of A in Q*, let alone of A in any subsystem of A. Indeed our next main step will be to show that under certain circumstances, it is not possible to obtain such improvements.[166]

The 'next main step' begins with the important notion of a reflexive theory.

> (Fef 5.7) Let $A = \langle A, K \rangle$, $K_0 \subseteq K$. A is said to be *reflexive* if for each finite $F \subseteq A$,
>
> $A \vdash \underline{Con}_{[F]}$.

Consequently, A is *reflexive* just in case

for each n, $A \vdash \underline{Con}_{[A/n]}$.

This definition is not empty, since Mostowski showed that

> (Fef 5.8)
>
> (i) PA is reflexive
>
> (ii) Every consistent extension A of PA, with the same constants as PA, is reflexive.[167]

---

[166] Feferman [1960] p57.
[167] See Mostowski [1952].

(The generalized version of the First Incompleteness Theorem given above as (Gödel 1) is a consequence of this result of Mostowski's, and (Fef 5.6).) Reflexive theories have the following property

> (**Fef 5.9**) Suppose that $A = \langle A, K \rangle$ is a consistent, reflexive axiom system with $\mathbf{PA} \subseteq A$. Suppose further that A is recursive. Then there is an φ^* which bi-numerates A in A for which
>
> $$\mathbf{PA} \vdash \underline{\mathrm{Con}}_{\varphi^*}.$$

Proof: Since A is recursive, there is some φ which binumerates A in **PA**. Define a formula φ^* in one free variable as follows:

$$\varphi^*(x) = \varphi(x) \wedge (\forall z)(z \leq x \rightarrow \underline{\mathrm{Con}}_{\varphi/z}) \wedge \underline{\mathrm{St}}_K(x).$$

If φ binumerates x, then clearly φ^* binumerates x also. But the second conjunct of the definition ensures that φ^* has additional properties. In particular, it ensures that the (sets of) axioms preceding x in some ω-ordering of axioms are consistent. Then, if $n \in A$, we have

$$(1)\ \mathbf{PA} \vdash \varphi(n) \wedge \underline{\mathrm{St}}_K(n)$$

whence by the hypothesis of the theorem

$$(2)\ A \vdash \varphi(n) \wedge \underline{\mathrm{St}}_K(n)$$

And since A is reflexive,

$$(3)\ A \vdash \underline{\mathrm{Con}}_{\varphi/0} \wedge \ldots \wedge \underline{\mathrm{Con}}_{\varphi/n}$$

whence

$$(4)\ A \vdash \varphi^*(n).$$

And if $n \notin A$, then

(5) $A \vdash \neg\varphi(n)$

whence by the construction of φ^*,

(6) $A \vdash \neg\varphi^*(n)$

and therefore φ^* also binumerates A in A.

The theorem is now obtained as follows. By earlier results,

(7) $M \vdash \underline{Con}_\varphi \leftrightarrow (\forall z)Con_{\varphi/z}$

whence

(8) $M \vdash \neg\underline{Con}_\varphi \rightarrow (\exists z)(\neg\underline{Con}_{\varphi/z} \wedge (\forall y)(y \leq z \leftrightarrow \underline{Con}_{\varphi/y}))$.

But then by construction

(9) $M \vdash \neg\underline{Con}_\varphi \rightarrow (\exists z)(\underline{Con}_{\varphi/z} \wedge (\forall x)(\varphi^*(x) \leftrightarrow \varphi(x) \wedge x \leq z))$

so

(10) $M \vdash \neg\underline{Con}_\varphi \rightarrow (\exists z)(\underline{Con}_{\varphi/z} \wedge (\forall x)(\underline{Pr}_{\varphi^*}(x) \leftrightarrow \underline{Pr}_{\varphi/z}(x))$

from which it follows that

(11) $M \vdash \neg\underline{Con}_\varphi \rightarrow \underline{Con}_{\varphi^*}$.

On the other hand, by the construction of φ^*,

(12) $M \vdash (\forall x)(\varphi^*(x) \leftrightarrow \underline{Fm}_K(x) \dashrightarrow \varphi(x))$

and thus by (Fef 4.10) above,

(13) $M \vdash \underline{Con}_\varphi \rightarrow \underline{Con}_{\alpha^*}$

Then from (11) and (13),

(14) $M \vdash \underline{Con}_{\alpha^*}$

As a corollary of this, we have

(Fef 5.10) There is a φ^* which binumerates the axioms of **PA** in **PA** for which

$$PA \vdash \underline{Con}_{\varphi^*}$$

The Feferman-consistency of **PA** can be proved in **PA**.

What has been shown here? Well, looking at the crucial definiendum of φ^* in the proof of **(Fef 5.9)**, we have seen that it says something like: 'x is an axiom of **A**, and each finitely axiomatized subsystem of **A** generated by axioms 'shorter' than x is consistent'. One (perhaps unfair) way of reading the consistency sentence formulated using φ^* is therefore this: *the largest consistent subsystem of A is consistent.* Since **A** (**PA**, as it might be) *is* consistent, the largest consistent subsystem of **A** is **A**, and this is therefore a consistency proof for **A**. What prevents the demonstration of the *consistency* of **PA** in **PA** is the fact that not: $PA \vdash \underline{Con}_{\varphi} \leftrightarrow \underline{Con}_{\alpha^*}$, for otherwise (by 14) $PA \vdash \underline{Con}_{\varphi}$, and **PA** would then be inconsistent.

The situation, then, is rather pretty. *Given* the consistency of **PA**, $\underline{Con}_{\varphi}$ and $\underline{Con}_{\varphi^*}$ are indeed equivalent - consistency and Feferman-consistency come to the same thing. But *given* the consistency of **PA**, that equivalence cannot be demonstrated in **PA**.

Although it is, I think, fundamentally fair, this somewhat facetious account of what Feferman has shown must not be allowed to obscure the philosophical interest of the above argument. For, in the ordinary way, one would think that an axiom schema is something like an expression type that provides one with a way of recognizing instances of that type as axioms. But how do we recognize an instance as an instance of that type? Feferman has shown that there is *a* way of 'recognizing' instances of the axiom schemata used in formalizations of arithmetic on which the consistency of arithmetic *can* be proved in arithmetic. To be sure, this is not the ordinary way. What is more, the result depends upon the consistency of arithmetic, since in the crucial definition of φ^* above, the

consistency of a set of arithmetic axioms A is assumed. This might make Feferman's argument of no use in a dispute with a skeptic about the consistency of arithmetic. But there are reasons for being interested in syntactic proofs of consistency other than the desire to combat skepticism.

However, it should not be thought that a proof of Feferman-consistency is of much help from the point of view of Hilbert's Programme. For as we have just seen, in order for some formula A of some Feferman-system of arithmetic T to count as an axiom of T, A must not only be an axiom in the ordinary sense - that is to say, it must not only pass the constraint on axiomhood (whatever it is) imposed by the predicate φ from which the deviant predicate φ^* is defined - it must also be consistent with the axioms prior to A in some ω-ordering of axioms of T. And since, by Church's Theorem, there is no effective procedure for determining the consistency of sets of quantificational formulas, T cannot be a formal system in the Hilbertian sense. There is no effective test for T-axiomhood.

However, further discussion of Feferman- and other deviant systems will take us away from the curious logical facts surrounding the Second Incompleteness Theorem towards the philosophical issues to which those facts give rise. It will be best now to tackle those issues directly.

Section Three: The Second Incompleteness Theorem and Hilbert's Program. What impact does the Second Incompleteness Theorem have on Hilbert's Programme? Towards the end of Gödel [1931], Gödel writes:

> The results of Section 2 [the proof of the First Incompleteness Theorem, that is] have a surprising consequence concerning a consistency proof for the system P (and its extensions), which can be stated as follows:
>
> Theorem XI. Let k be any recursive consistent class of FORMULAS; then the SENTENTIAL FORMULA stating that k is consistent is not k-PROVABLE; in particular, the consistency of P is not provable in P, provided P is consistent . . .. [168]

This is the first statement of the Second Incompleteness Theorem.

Gödel (1931) contains no proof of this theorem, although it does contain a proof sketch. Having given this sketch, Gödel continues:

---

[168] Gödel [1931], p193.

> I wish to note expressly that Theorem XI . . . [does] not contradict Hilbert's formalistic viewpoint. For this viewpoint presupposes only the existence of a consistency proof in which nothing but finitary means of proof is used, and it is conceivable that there exist finitary proofs that cannot be expressed in the formalism of P . . ..[169]

Much later, however, Gödel's views on the implications of Theorem XI for Hilbert's Programme changed. In a note appended to the above passage in 1963, he writes:

> In consequence of later advances, in particular of the fact that due to A.M. Turing's work a precise and unquestionably adequate definition of the general notion of formal system can now be given, a completely general version of Theorems VI [the First Incompleteness Theorem] and XI is now possible. That is, it can be proved rigorously that in every consistent formal system that contains a certain amount of finitary number theory, there exist undecidable arithmetic propositions and that, moreover, the consistency of any such system cannot be proved in the system.[170]

Although he does not explicitly say so, it seems clear enough that this footnote withdraws the reservations expressed in the original text. By 1963, then, Gödel was of the opinion that the 'completely general' versions of theorems VI and XI do indeed 'contradict Hilbert's formalistic viewpoint'.

It is important, though, that Gödel's reservations were withdrawn only after he became convinced of the possibility of suitable 'completely general' versions of the incompleteness theorems. In Gödel's view, then, the difficulty for Hilbert's Programme comes, not from the specific results proved in Gödel (1931) themselves, but rather from the 'completely general' versions of them.

And that is obviously right. There is no difficulty for Hilbert's Programme attendant upon the fact that some particular formal system - Gödel's system **P**, as it might be - cannot prove its own consistency. If there is to be so much as the appearance of a difficulty here, it must lie in the possibility of generalizing Gödel's results over a large class of systems, including in particular the systems, mentioned in **Chapter One**, involved in carrying out the Conservation Programme for analysis.

However, we have just seen at some considerable length that the two incompleteness theorems differ very markedly with respect to the possibility of straightforward, natural generalizations to arbitrary formal systems. The transition from the specific theorem

---

169 Gödel [op cit], p195.
170 Gödel [loc cit].

proved in Gödel [1931] as Theorem VI to (**Gödel 1**) is relatively straightforward and philosophically uncontroversial: but not so with generalizations of Theorem XI.

To see the worry to which this gives rise, consider the following line of thought. The formulation of the First Incompleteness Theorem given above in ordinary English as (**Gödel 1**) readily suggests a corresponding formalized and clearly mathematical sentence. By this, I mean that anyone minimally experienced in these matters will feel confident that there is a formal sentence of the kind typical of formalized mathematical theories which is naturally translated into English as (**Gödel 1**). It is easy to feel convinced, then, that (**Gödel 1**) is associated with a genuinely mathematical claim in this sense.

Things are otherwise with the Second Incompleteness Theorem. A typical informal version of this might be

> (**G2***) No complete, consistent, axiomatizable extension of **Q** can prove its own consistency.

(**G2***) does not so readily suggest a corresponding, genuinely mathematical sentence. To begin with, (**G2***) is modal: it speaks of what cannot be done, not of what does not exist. This defect is remediable, and a first attempt at removing it is apt to result in something more nearly equivalent to Gödel's own formulation, such as

> (**G2****) If **T** is any complete, consistent, axiomatizable extension of **Q**, then the sentence $\neg Bew(\ulcorner 0=1 \urcorner)$ is not provable in **T**.

But there are two problems about this, In the first place, the ghost of modality lingers on in the phrase 'not provable'. Once again, a remedy comes easily to hand, and the result of applying it will be

> (**G2**) If **T** is any complete, consistent, axiomatizable extension of **Q**, then there exists no proof in **T** of the sentence $\neg Bew(\ulcorner 0=1 \urcorner)$.

And this is now close to (**Gödel 2**). Secondly, though, there is the problem of the predicate denoted in (**G2**) by the expression '$Bew(\ulcorner 0 = 1 \urcorner)$'. This of course is to be a provability predicate for **T** - a predicate that is not only defined as $Bew(y) = (\exists x)Pr(x, y)$ for *some* predicate Pr which numeralwise expresses the provability relation of **T**, but is also so defined with respect to a predicate Pr which *also* satisfies the derivability conditions

(DER 1) - (DER 3). If (G2) is to have the import that the non-mathematical looking but philosophically interesting (G*) has - if, that is, the mathematical result (G2) is to be given as evidence for the philosophical claim (G*) - then something will have to be said about the restriction of our attention to predicates which not only give numerically correct representations of the proof relation of a theory, but also satisfy the derivability conditions. And the task is the more urgent, of course, since a liberalization of this policy will allow for the consideration of Feferman-style numerically correct representations of the proof relation of a theory with respect to which (G*) fails.

Detlefsen argues that no justification of the Derivability Conditions can be given.[171] Specifically, he rejects two particular justifications of the derivability conditions defended in the literature by Mostowski, and by Kreisel and Takeuti respectively.[172] Let us look at these arguments.

Mostowski considers the question, What are the demands that ought to be satisfied by a satisfactory arithmetization of the syntax of some system T? He writes:

> . . . given, on the one hand, a set X of integers (or pairs, triples etc.) and, on the other hand, a formal language. We are looking for the best possible definition of X in T, i.e., for a definition which makes, of all the intuitively true formulae involving X, as many as possible provable in T.[173]

In the proof of the First Incompleteness Theorem, for example, we have a set X = {<n, m>: n is the code of a proof in T of the formula with code m}, and we require that the representing formula of X numeralwise express P(x, y). This makes all instances of 'F is a proof in T of G' and 'F is not a proof in T of H' theorems of the arithmetized metamathematics of T. But then, Mostowski notes, not all truths concerning the provability relation have been captured by the arithmetized metamathematics, for in addition to all these particular truths concerning provability, there are also the general truths concerning provability. And we have just seen at length that not all formulas that numeralwise express the provability relation do an equally good job of capturing what we 'intuitively' regard as general truths concerning provability.

---

171 See Detlefsen [1986], esp. chapters Three and Four.
172 See Mostowski [1966], esp. pp25 ff, and Kreisel and Takeuti [1974].
173 Mostowski [op cit] p25.

This, of course, is where the Derivability Conditions come into consideration. Mostowski reasons, not implausibly given the facts discussed above, that the work of Bernays and Feferman shows that certain metamathematical tasks - such as those that feature prominently in Hilbert's Programme - call for a closeness of fit between an arithmetical representation of a metamathematical concept and the metamathematical concept represented that goes beyond mere extensional correctness. And his suggestion is, then, that adequate closeness of fit is to be determined by the number of intuitive truths regarding the metamathematical concept yielded by the arithmetization. The suggestion is, then, that the conditions **(DER 1)** - **(DER 3)** will capture a greater number of intuitive truths concerning the provability relation than any alternative.

The problem here, according to Detlefsen, is that that no arithmetization can capture *all* the intuitive truths concerning this notion. The undecidable formula '17 Gen r', and the consistency sentence constructed from it, provide the obvious examples. But if we cannot capture all of the intuitive truths we might be interested in, why should we be so interested in capturing *most* of them? Detlefsen writes:

> Evidently, if representation of this notion can occur without capturing all such truths, then only certain truths are crucial to the ability of a formula to express it. And if that is the case, then the important thing is to capture *those* truths, and not to capture as many truths as is possible.[174]

The question is, though, which truths should we be interested in? Detlefsen points out that a Hilbertian will be particularly interested in capturing truths about *unprovability* in T and the closely related notion of consistency. But here the Derivability Conditions fare very badly, for they make it impossible to represent *any* truths regarding unprovability in T as theorems of the arithmetized metamathematics of T, on pain of showing T inconsistent (since any statement of underivability is tantamount to a statement of consistency). Mostowski's proposal, Detlefsen therefore contends, does not offer a very convincing defence of the Derivability Conditions as devices for excluding Feferman-style 'provability' predicates with respect to which arithmetical consistency is provable.

However, effective as Detlefsen's argument is against the letter of the Mostowski proposal, it does not seem to me so effective against the spirit. To see why, we must reflect more generally on the nature of arithmetization, and indeed on formalization as a whole.

---

[174] Detlefsen [1986] p104.

We start out with standard English, or a mildly mathematical extension thereof. Standard English provides us with expressions denoting such things as variables, of the left bracket, or the comma. I have just used them. I take it that it is not required of me to say what Standard English having this capacity amounts to. That is a fascinating question, but it is not a question the philosophy of mathematics ought to be expected to answer. Standard English also has predicates satisfied by formulas of specified language, such as 'x is a formula of the language QC', or by proofs in a specified formal system, such as 'y is a proof in QC'. Standard English allows us to discuss the symbols of **PRA** and their permissible combinations.

George Boolos calls the part of Standard English in which we discuss such things, the *language of Syntax*, and the informal theories we construct in that language, *Syntax*. Let us adopt Boolos's usage. And for the sake of simplicity, let us restrict Syntax to the study of the language of **PRA**.

Syntax is, then, the general theory of well-formedness of certain (types of) strings of symbols of **PRA**. As a special case of well-formedness of strings of symbols, Syntax studies *derivations*. More precisely, Syntax tells us under what conditions a string of well-formed formulas - this can be taken to be a certain kind of sequence - constitutes a *derivation* - when each *element* of the *sequence* is a *wff* which is either an *axiom*, or the result of performing certain *syntactic operations* on *antecedent wffs*. In each case, the italicized expressions are defined in Syntax.

Arithmetization is simply a tool to facilitate the study of Syntax. In 1931, when Gödel first demonstrated the powers of this tool, very little was known about Syntax, but a great deal was known about the elementary theory of numbers. Arithmetization enabled the study of Syntax to draw upon this rich fund of resources. For example, little was then known about concatenation in general, and in particular about concatenation under the restrictions imposed by arbitrary transformation rules. Under arithmetization, it can be seen that facts about concatenation, both in general and under restrictions, can be formally expressed as facts about factorization of natural numbers, by regarding strings as numerals, and syntactic operations as arithmetic operations. Since we then knew a great deal about congruences, factors, remainders etc., this systematic correspondence enabled us to understand Syntax much better.

But there is no reason other than convenience for Syntax to draw upon the resources of arithmetic via arithmetization. For Gödel and Quine between them have shown that elementary number theory and Syntax are mathematically equivalent. Exactly the same formal system can be indifferently interpreted as Syntax, or as number theory.

In *this* sense, then, there is rather less to arithmetization than meets the eye. The incompleteness theorems, for example, are standardly proved via arithmetization. But they need not be. If we use 'Syntax' for the formal system that stands in the same relation to Syntax as PRA stands to PRA, then the incompleteness of Syntax can be demonstrated by Syntax, in Syntax.[175] The detour through arithmetization makes things much nicer, much easier to follow, much more elegant; but it is not strictly necessary.

Mostowski, therefore, should not be making a point about arithmetization: his point should be a point about formalization in general. For there is nothing special about arithmetization as a means of bringing some naively given theory under proper scientific control. Arithmetization is a tool, no more mysterious in its formalizing efficacy, no more controversial in applications than the familiar first-order predicate calculus. And we to not resort to the tool of formalization in order to capture some, or most, of the 'intuitive' truths given in the unformalized theory. We expect a formalization to capture *all* of them, and if a formalization does not do so, then we regard that either as evidence that the formalization is inadequate, or that our intuitions need careful scrutiny. In the case of naive semantics, formalization quickly reveals that our intuitions need careful scrutiny, whereas the case of first order formalizations of PA might be thought to support the alternative reaction. Typically, there will be a consensus as to which of these alternative to adopt, and there will also be a consensus as to which intuitions in particular need more careful scrutiny. If this were not so, the naive theory would have to be junked - as astrology was junked - as scientifically unworkable.

In any formalization of a naive theory, there will be considerable leeway at the outset of the enterprise. Syntax, for example, can treat wffs as strings, or as trees. Strings and trees are different kinds of mathematical structure, and so far as I can see, there is no reason to believe that Syntax formalized as a theory of strings need be mathematically equivalent to Syntax formalized as a theory of trees. If that is right, perhaps we shall have to decide

[175] I know of no detailed example of such a demonstration in the literature, but Smullyan [1961] gives an outline of how this could be done.

between $\mathbf{Syntax}_{\mathbf{strings}}$ and $\mathbf{Syntax}_{\mathbf{trees}}$ as optimal formalizations of Syntax. But notice that the reason that we might be able to make sense of there being a rational *decision* here suggests strongly that, after the initial parameters of the formalization have been fixed, the enterprise proceeds under pretty stringent constraints. Flexibility is strictly limited. Once you have committed yourself to formalization of Syntax in the fashion of $\mathbf{Syntax}_{\mathbf{strings}}$, the properties of strings and the content of Syntax combine to greatly reduce your freedom of choice as to how the formalization should proceed. Once you have decided how to treat the primitive notions of Syntax - the primitive expressions, and the primitive operation of concatenation - then there really is an obvious natural formalization of the compounds of those primitives, up to and including derivations. For a derivation in Syntax is sequence (or perhaps a tree) meeting such and such conditions, themselves formalizable in a tightly constrained way. Given those conditions, the natural formalization in **Syntax** of the notion of a *derivation* just drops out, as we saw above in the discussion of Feferman.

And (to repeat the point I made above), if you do this in the natural way for Syntax, then the predicate you introduce to play the role of Bew(y) in proving the incompleteness theorems *will in fact satisfy* the derivability conditions. They do not have to be 'imposed', and they are not 'constraints'. They simply characterize the standard notion of derivation, and this can be mathematically demonstrated once that notion is formalized in the obvious way. Mostowski, then, is not making a controversial point about the conditions that should be met by an adequate arithmetization. He is making - well, mis-stating would be strictly more accurate - a perfectly mundane and uncontroversial claim about formalization. No-one should dispute this claim, least of all a Hilbertian.

What this suggests to me, therefore, is that what Detlefsen is really concerned about is not the adequacy of the *formalization* of the standard notion of proof, but rather the adequacy of the standard notion of proof itself. Rather than denying that the standard notion of proof complies with the Derivability Conditions, Detlefsen's real target is the standard notion of proof. Consideration of his response to the Kreisel-Takeuti proposal seems to me to confirm that this is what is really animating his discussion.

Kreisel and Takeuti write

> . . . the usual conditions on systems [they in fact have in mind the original Bernays derivability conditions, but we may take these to be (DER 1) - (DER 3)] . . . are necessary if a formalization of mathematical reasoning is to be

> adequate for Hilbert's programme . . . Let us spell out two adequacy conditions on a system F:
>
> (a) Demonstrable completeness w.r.t. Σ_1 formulae is needed to assure us that elementary mathematics (with a constructive existential quantifier) can be reproduced in F at all . . .
>
> (b) Demonstrable closure under cut (and in the quantifier free case also under substitution) is also needed because cut is constantly used in mathematics. Realistically speaking, a (meta)mathematical *proof* of such closure is needed and not a case study of mathematical texts because cut - like most logical inferences - is often used without being mentioned; in contrast, for example, to the use of mathematical axioms.[176]

Demonstrable completeness w.r.t. Σ_1-formulae guarantees (**DER 1**) and its formalization (**DER 3**), whilst demonstrable closure under cut guarantees (**DER 2**). The Kreisel-Takeuti proposal, then, is that the Derivability Conditions must apply to any provability predicate Bew(y) which is adequate to capture ordinary, informal mathematical reasoning. What is more, a metamathematical demonstration of this is the only 'realistic' assurance that we have that the adequacy conditions are in fact met. Detlefsen summarizes this position thus:

> . . . the Kreisel-Takeuti position consists of both a substantive and a procedural claim. The substantive claim is that if T is to be an adequate formal codification of informal mathematical practice, then it must be both Σ_1 complete (so that it captures elementary mathematics) and closed under modus ponens (so that it is sure to capture the logical technique of informal classical mathematics). And the procedural claim is that the only practical, or at least the epistemologically optimal, way of coming to know that the above substantive conditions are met is via a T-codifiable metamathematical proof of them.[177]

In his response to the Kreisel-Takeuti proposal, Detlefsen then argues that it appeals to an unreasonable *idealization* of informal mathematical practice. He writes:

> . . . if what Kreisel and Takeuti mean by mathematical practice is the rounding out of actual historical practice to conditions of perfect rationality and information, then perhaps they are right to say that T cannot hope to codify mathematical practice unless it is closed under cut and arithmetically complete. . . . But should the appropriateness of the Derivability Conditions be judged from such a standpoint? More specifically, should the appropriateness of the Derivability 'Conditions, as constraints governing the Hilbertian's choice of formalizations of ideal reasoning, be judged from such a standpoint? We think not, since though the ability of his formalisms to capture informal mathematical practice is certainly a concern of the Hilbertian's it is by no means his only, or even his dominant, concern. . . . . Chief among the factors which draw the Hilbertian away from excessive idealization are his need to obtain a soundness proof for his formalizations of ideal reasoning, and his natural lack of concern for

---

[176] Kreisel and Takeuti [1974] pp34-35.

[177] Detlefsen [1986] p115-116.

> ideal reasoning which is too long or complex to be of any likely benefit . . [178]

But what can this mean? Clearly, the belief-set of the average (or indeed the extraordinary) mathematician is not closed under cut, and is not Σ_1 complete. Indeed, even if we pool our collective mathematical wisdom across the ages, it is neither closed under cut, and is not Σ_1 complete. But so what?

Clearly, the major problem we face in discussing this dispute lies in getting some sort of a grip on this notion of 'informal mathematical practice'. Detlefsen suggests understanding this as 'the totality of assertions and justifications that have gained the popular acceptance of the historically given community of mathematicians'.[179] But that does not seem to me to be the natural, or the best, proposal. The natural proposal, I think, is to look for the *norm* that governs that 'popular acceptance' of assertions and justifications.

But it seems to me that there is no mystery about what that norm is, at least in bold outline. Justification in mathematics consists of proof, authority in mathematics is authority with respect to the existence of proofs, and assertion is assertion that some proof exists. What is a proof? So far as the ordinary practice of mathematicians is concerned, I know of no reason to doubt that it is just what the formal logicians tell us it is - a finite array of formulas, each of which is either an axiom, or follows from previous formulas . . . etc.

But, it will be objected, if this is what proofs are, then mathematics as it is actually practised contains no proofs. Quite right. No-one ever displays derivations from axiom systems in doing real mathematics: one can get through a graduate education in mathematics without ever seeing a proof.[180] Rather, real mathematics consists of arguments, more or less informal, *one primary purpose of which is to make it sufficiently clear that a proof exists.* One can, in this sense, 'give' a proof without writing it down.

What it takes to fulfil this purpose, of course, varies according to such factors as the intended audience, and also the familiarity of the branch of mathematics in which one is

---

178 Detlefsen [op cit] pp116-117.

179 Detlefsen [op cit] p116.

180 From the ordinary mathematics texts on my bookshelves, I pick (pretty much at random) Apostol's two volume *Calculus* - an unusually theoretical undergraduate text - and a quick scan reveals not a single instance of a proof. Selecting now from the logic texts on my bookshelves, a quick scan of Monk [1976] reveals one proof - on page 118. Perhaps there are a couple more.

working. In so central a branch of mathematics as number theory, for example, the degree of informality is normally very high. In areas in which the subject matter is less familiar, and in particular in areas in which is or recently has been controversy, the standards of informality will normally be much lower. But so far as I can judge from the outside, at least, if there is ever real doubt that a genuine proof of some assertion can be constructed from an informal argument justifying that assertion, the standard practice of mathematicians is to improve, and if necessary complicate the argument until the existence or non-existence of a genuine proof becomes obvious to all competent judges. It is in this sense that the norm of formal proof governs the ordinary practice of mathematics.

The idealization that is implicit in the Kreisel-Takeuti proposal, then, is not an idealization *of* ordinary mathematical practice. Rather, it is the idealization that is commonly accepted *in ordinary mathematics* as providing the norm governing ordinary mathematical practice. Kreisel and Takeuti are not 'rounding out . . actual historical practice to conditions of perfect rationality and information', for they are not making claims about the belief-set of some idealized mathematician. But then, one might of course take issue with the appropriateness of this norm. One might think that the Rosser proof predicate, say, provided a more appropriate norm for mathematical practice. Disputes of this general kind have been very common in the history of mathematics - Brouwer's intuitionist alternative to classical mathematics, for example, is grounded in an attack on the appropriateness of classical logic as the basis for mathematical practice. Detlefsen, in his advocacy of some alternative notion of proof (Rosser proof, as it might be) according to which the consistency of arithmetic is finitistically provable, has no need to dispute that closure under cut and demonstrable Σ_1 completeness characterize the standardly accepted notion of provability. He needs to dispute the appropriateness of that notion.

Both in his response to Mostowski, and to Kreisel and Takeuti, therefore, I think that Detlefsen confuses two quite different issues. The first issue concerns the conditions that the standardly accepted notion of mathematical proof in fact satisfies. *That* notion of proof satisfies the Derivability Conditions, and so far as I can see, nothing Detlefsen has to say ought to incline anyone to doubt that the consistency of **PRA** cannot be proved in **PRA**, in *that* sense of 'proved'. The other issue, though, concerns the notion of proof that Detlefsen's 'Hilbertian' instrumentalist ought to accept. Perhaps the appropriate notion is that of *Roof* (for Rosser-proof), rather than proof. Detlefsen then has a perfectly sensible, interesting project to recommend to us - the project of discovering the properties of Roovability, and in particular, the project of discovering finitary Roofs of consistency. Of

course, we shall as Hilbertians require that the notion of Roof is axiomatized, for at the moment its characterization is entirely parasitic upon the well-understood notion of proof. Once we have a Roof-predicate axiomatized independently of the standard proof predicate, we shall be able to encode it and investigate its metamathematical properties.

But we must then wonder what Detlefsen will have to say about those theorems of standard mathematics that cannot be Rooved, although they can be proved, and those theorems that can be proved, but not Rooved. For clearly, the two notions cannot be co-extensive. If Detlefsen's instrumentalism has the consequence that the unRoovable fragment of classical mathematics is to be abandoned, or in any way downgraded in the interests of the Roovable fragment, then he is engaged in exactly the kind of mathematically revisionary philosophical assault on ordinary mathematics that Hilbert's Programme was intended to forestall.

And there is much in his book that suggests that this is what he has in mind. In particular, the kind of instrumentalism he advocates appears to be animated by the all-too-familiar kind of epistemology-based worries about Platonism, for he makes much of the need for a 'justification' of infinitistic mathematics that does not require literal belief in its deliverances. The justification he tries to give, as we saw at the close of **Chapter One**, has as its central theses that infinitistic mathematics is a 'reliable' and 'perspicacious' extension of finitary mathematics. These are technical terms in Detlefsen: an extension of finitary mathematics is 'reliable' only if it proves 'finitarily true' finitary theorems only, and 'perspicacious' only if it proves all (or at least lots) of the finitary truths we are interested in more readily that finitary mathematics does. And as I said above, there is real point of contact with Hilbert here, since this is the Conservation Programme in other words. Detlefsen's instrumentalism, then, really is seriously challenged by the Incompleteness Theorems, which show that ordinary, infinitistic mathematics is either unreliable, or non-perspicacious. But if Detlefsen proposes to redraw the map of ordinary mathematics in order to get reliability and perspicacity within the confines of the notion of Roovability, he has clearly abandoned the Hilbert's project of defending the ordinary practice of mathematics. For Detlefsen has nothing to say about the justification of the parts of ordinary mathematics which cannot be accommodated within the confines of this instrumentalism.

The Hilbertian philosophy of mathematics I sketched in **Chapter One** and **Chapter Two** does not have this deficiency, since it separates issues concerning ontology and issues

concerning justifiability. The justification of mathematics is in general a matter for the internal practice of mathematics, not for the philosophy of mathematics. In the special case of finitary mathematics, of course, a different and more philosophical kind of 'justification' can be given, which rests upon a claim about the special status of mathematics in any theory of representational thought. In part as a consequence of this, my version of Hilbertian philosophy of mathematics is committed to mathematical objects, in theories that are finitarily provably conservative over finitary mathematics. But it is not compelled to abandon literal belief in the infinitistic parts of mathematics beyond conservative extensions of the finitary, nor to deny that the well-attested theorems of transfinite set theory are true. For my Hilbertian is no instrumentalist: she treats the truth predicate disquotationally, and regards the internal practice of mathematics as in general perfectly adequate to confer meaning upon mathematical statements, and thus legitimate literal belief in what they say. My Hilbertian, then, can abandon the Conservation Program with comparative equanimity, and rest content with the standard understanding of the import of the incompleteness theorems.

Detlefsen's instrumentalist, then, does not seem to me to provide a plausible philosophical basis for anything like Hilbert's Programme. Hilbert was no revisionist, and there is no reason to think that he would have abandoned classical mathematics in the interests of a viable version of the consistency programme. Nor can I see any reason why Hilbert should be interested in anything other that a *proof* of consistency. For Hilbert's interest in consistency was above all focussed upon the problem of the elimination of 'ideal elements' from proofs of theorems of finitary character. That project accepts classical mathematics as it is, and seeks to show that the classical quantifier can always be eliminated from formalized proofs of finitary theorems.

But it cannot be, and that is why Hilbert's Programme, in its original, fully general form, fails. What is more, we can see that it fails without becoming embroiled in the surprising properties of the notion of provability in a formal system, or indeed with any controversial issues in the interpretation of formal theories. For the fully 'extensional' First Incompleteness Theorem will suffice to force this conclusion upon us, as I shall now show.

Section Four: The First Incompleteness Theorem and the Master Argument. Recall that the Master Argument aimed to demonstrate the eliminability of ideal elements

from proofs of finitary theorems.[181] In this section, I shall give an account of some work by Kripke and Goldfarb that shows the existence of a version of the First Incompleteness Theorem which provides a direct demonstration that the Master Argument cannot be carried out within the constraints of finitary mathematics. The qualification is important: the Master Argument *can* be carried out, but the weakest possible resources required for its successful execution go just beyond the bounds of the finitary. This has of course been known since the original Genzen proof of the consistency of arithmetic of 1936 (see Genzen [1936]). What is perhaps less widely known is the connection between Genzen's discoveries and the incompleteness theorems.

You will also recall from **Chapter One** Hilbert's insight that it is 'through the quantifier' that the infinite enters mathematics. The basic strategy for proving consistency pursued by Hilbert and his collaborators was therefore to seek out an algorithm for the elimination of occurrences of quantifiers from proofs. Like Skolem, but independently of him, Hilbert took the view that an existentially quantified variable may be thought of as a choice function. In a formula $(\forall x)(\exists y)F(x, y)$, e.g., the value(s) of the dependent variable y required for the truth of the formula may be thought of as the value(s) of a function which, given an arbitrary x, chooses a y such that F(x, y). The sense of $(\forall x)(\exists y)F(x, y)$, then, is roughly that of $(\forall x)F(x, f(x))$, where f(x) is the appropriate choice function.[182]

The demands upon such a choice function are apparently very strong, since it must not only provide values for all the (possibly infinitely many) values of the universally quantified variable (or variables) upon which it depends, but also choose values in such a way as to guarantee truth, if that is possible. Given the possibility of extensive nesting of quantifiers, it will not in general be possible to constructively verify the truth of the outcome of the operation of the choice function. This is of course why the quantifier counts as an ideal element, in Hilbertian terms.

We saw in **Chapter One** the basic source of Hilbert's hopes for a way around this difficulty. With arithmetic formalized in the ε-calculus, it will be possible to convert any formula, and thus any proof, into an equivalent formula or proof in which all quantifiers have been replaced by ε-terms. Since any particular proof can only invoke finitely many

---

[181] This section is very heavily indebted to the work of Warren Goldfarb - see especially Goldfarb [1990].

[182] This rough claim can be made precise. The *Skolemized form* of a formula of the predicate calculus is the result of replacing all occurrences of existentially quantified variables in the manner suggested by the example. It can then be shown that a formula is contradictory iff its Skolemized form is contradictory.

axioms, and can apply inference rules only finitely often, Hilbert thought that any particular proof need only draw upon a finitistic fraction of the full infinitary power of the choice function(s) associated with the ε-terms. Oversimplifying drastically, if a proof contains only n occurrences of ε-terms, one may think of the algorithm for eliminating ε-terms as a systematic search through n-tuples of natural numbers. If there is an assignment of values to variables that validates the proof, Hilbert thought, it must be located within the bounds of a finite search. Goldfarb describes this procedure as a search for *finitary approximations* to the genuine, infinitary choice functions required. One can start with the approximations to the real choice functions that are zero everywhere, then check if any of the constraints imposed by the formulas in the derivation are violated and halt if they are not; otherwise, begin some systematic process of varying the previous assignments and repeat.[183]

This is the strategy pursued by Ackermann in his proof of the consistency of first order arithmetic.[184] The difficulty it faces is rather obvious to us now, but Hilbert and his school were apparently oblivious to it until the discovery of the First Incompleteness Theorem in 1931. In order to carry out the Master Argument by the substitution method, it will have to be shown that the process of assigning values to ε-terms until a validating assignment is produced always terminates after a finite search. This is going to have to be argued by induction (in fact, a double induction, on the complexity of formulas and the length of proofs), and that in turn demands some notion of normal form for formulas and proofs. Satisfactory notions are not hard to find. One possibility for the normal form of formulas is prenex normal form, and this makes vivid the problem that will eventually drive us beyond the bounds of the finitary. For we will have to be able to deal with quantificational prefixes of arbitrary complexity.

Now, as we saw in the simple example of the Skolemization of $(\forall x)(\exists y)F(x, y)$ to $(\forall x)F(x, f(x))$, the function that replaces the dependent existentially quantified variable must be keyed to the values of the variable upon which the existentially quantified variable depends. Given a more complex matrix, this induces complex relations of functional dependence, since the constraints imposed upon the approximation to the choice function replacing one existential quantifier will depend upon the constraints imposed upon other approximations to choice functions in whose scopes the original lies. Consequently, in the induction argument to the conclusion that the substitution procedure always terminates, the

---

[183] This procedure is known as the substitution method. The standard study is Tait [1965].
[184] Ackermann [1940]. There is a readable version of Ackermann's proof in Wang [1963].

induction has to cope with structures of greater ordinal complexity than the natural numbers. In Genzen [1936], and in somewhat greater detail in Scanlon [1973] (which draws upon the work by Herbrand discussed below), it is shown that the argument in fact requires induction up to ordinal ε_0 - the least ordinal with respect to which transfinite induction cannot be reduced to ordinary induction in a finitary way, and in this sense it has been shown that the minimal resources required for the proof that the substitution method terminates *just exceed* the bounds of the finitary.

This much is widely known. What is less widely known, perhaps, is the intimate relation between these facts and the incompleteness theorem. Goldfarb, following unpublished work by Kripke, has established this connection as a consequence of the profound investigations in proof theory undertaken by Herbrand in his doctoral thesis.[185] I shall now give an account of Goldfarb's argument.

In his PhD thesis, Herbrand applies Hilbert's ideas for finding finitary approximations to the quantifiers to the first order predicate calculus. Rather than speaking, as I have, of choice functions (and thus of a domain of objects in some intended model), Herbrand speaks of choices of *new function signs* to replace the quantifiers. The basic idea is to show that any quantificational formula can be effectively reduced to a countable set of formulas free of quantifiers and variables, i.e. to what is essentially a set of formulas *of the propositional calculus*. The collection of quantifier-free formulas corresponding to some formula $(Qx_i)F$ is called the *Herbrand expansion* of F (denoted by E(F)). In this way, apparently quantificational facts about domains of objects are analysed as *truth functional facts* about their expansions.

Assume now we are working in a formulation QC of the first order predicate calculus with $\exists$, $\neg$, $\vee$, and $\wedge$ primitive, and $\forall$, $\leftrightarrow$, and $\rightarrow$ introduced as informal abbreviations in the usual way. A *rectified* formula of QC is a formula of QC in which (i) no variable occurs both bound and free, and (ii) there is at most one occurrence of a quantifier with any given variable. (Every formula of QC has a rectified equivalent formula in QC.)

Let F be a rectified sentence of QC. An occurrence of $\forall$ is *positive* (in F) if that occurrence of $\forall$ lies in the scope of an even number of occurrences of $\neg$ in F, and *negative*

---

[185] According to Goldfarb [1990], Kripke's proof dates from 1978. There is a sketch of it in part 7 of Kochen and Kripke [1981]. Herbrand's doctoral thesis is Herbrand [1930]. The crucial fifth chapter is reprinted in an English translation in van Heijenoort [1967] pp524-581.

otherwise. An occurrence of $\exists$ (in F) is called *positive* (in F) if that occurrence of $\exists$ lies in the scope of an odd number of occurrences of $\neg$ in F, and *negative* otherwise.

[Example: if

$$F = (\forall x)(F(x) \vee (\exists y)\neg(\exists z)(G(y, x, z) \vee (\forall w)\neg G(w, x, z)))$$

then x and z are bound by positive occurrences of quantifiers, and y and w are bound by negative occurrences of quantifiers.]

We now need to define the *functional form* of a rectified formula F. For each negative occurrence of a quantifier Q in F, let x be the variable bound by that occurrence of Q, let G be the scope of that occurrence of Q, and let $y_1, \ldots, y_n$ be in order the variables bound by positive occurrences of quantifiers in whose scopes QxG occurs. Choose a new function sign f_x not used elsewhere in F, and substitute $G[x/f_x(y_1, \ldots, y_n)]$ for QxG. Repeat this procedure until no negative occurrences of quantifiers remain. The result is the *functional form* of F (it is unique up to the choice of function signs). The new function signs are the *indicial function signs*.

[Example continued: the functional form of F above is

$$(\forall x)(F(x) \vee \neg(\exists z)(G(f_x(x), x,z) \vee \neg G(f_w(x, z)), x, z)))$$

and the indicial function signs are $f_x(x)$ and $f_w(x, z)$).]

What we have done is simply extend the notion of Skolemization to cover non-prenex formulas. Now, let F^- be the *matrix* of the functional form of F.

[Example continued: with F as before, F^- is

$$(F(x) \vee \neg(G(f_x(x), x,z) \vee \neg G(f_w(x, z)), x, z)))]$$

We define the *height* of a closed term inductively: constants are of height 0, and the height of $f(x_1, \ldots, x_n)$ $(n > 0)$ is one greater than the maximum of the heights of $x_1, \ldots, x_n$. We can now define inductively the *Herbrand universe of height p associated with F*, denoted by D(F, p). We include in D(F, p) an initial constant c (zero-place function),

which is not called a indicial function but which is included in D(F, p) for each F, p (this is just to allow the construction of D(F, p) to begin in case F^- contains no constant). Then, if $t_1, \ldots, t_n \in$ D(F, p) (for $n \leq p$) and f is any n-place function sign (not necessarily indicial) occurring in F^-, then $f(t_1, \ldots, t_n) \in$ D(F, p). D(F, p) is then the finite set of terms of height $\leq$ p constructed from the function signs of L(QC) and the indicial functions that appear in F^-, together with the initial constant c.

[Example continued: with F^- as above, D(F, p) would begin thus

$$\{c, f_x(c), f_w(c, c), f_x(f_x(c)), f_x(f_w(c, c)), f_w(f_x(c), c) \ldots\}$$

and would halt at the bound on the height of terms imposed by p.]

Finally, we can define the *Herbrand expansion of F of height p* , E(F, p) as the conjunction of all instances of F^- over D(F, p) - the conjunction of all sentences obtained from F^- by substituting terms from D(F, p) for its free variables (that is, the variables that were bound by positively occurring quantifiers in the functional form of F).

[Example concluded: the Herbrand expansion of height p of the formula F above might begin

$$\{(F(c) \vee \neg(G(f_x(c), c, c) \vee \neg G(f_w(c, c)), c, c))) \wedge$$
$$(F(f_x(c)) \vee \neg(G(f_x(c), c, c) \vee \neg G(f_w(c, c)), c, c))) \wedge \ldots\}]$$

We can now state *Herbrand's Theorem* :

> **(Herb)** A sentence F is derivable in any standard axiomatic system for **QC** iff E($\neg$F, p) is truth functionally unsatisfiable for some p.

For the proof, which is reasonably straightforward, see Herbrand [1971], or e.g. Andrews [1986] section 35. It is perhaps plausible from the above description, but in any case can be proved, that one great merit of **(Herb)** from the point of view of Hilbert's Programme is its effectiveness. Given a derivation of F, we can compute a bound on the search for an

unsatisfiable expansion of $\neg F$.[186] This makes (**Herb**) a remarkable confirmation of Hilbert's insight that any given derivation will only draw upon a finite portion of the full infinitistic power of the quantifier - so far as QC is concerned.

But (**Herb**) has other virtues for the Hilbertian, since it speaks directly to the question of consistency. A first order system T is consistent iff each conjunction of the axioms of T is quantificationally irrefutable, iff (by (**Herb**)) $E(\wedge(Ax_T), p)$ is truth functionally satisfiable for every $\wedge(Ax_T)$ and every p. The gain for Hilbert, now, is that *only quantifiers occurring in the axioms need to be considered in proving consistency.* As Goldfarb says:

> Everything becomes a matter of the axioms: it is only to the Herbrand function signs arising from them that a Hilbert-style evaluation procedure need be applied. This makes the project resemble more closely the intuitive (nonconstructive) one of devising a model for the axioms. But whereas the latter task requires real choice functions, functions that render every expansion E(F, p) true simultaneously, the proof-theoretic task is rather to find for each F and each p functions that work for E(F, p). The values chosen in connection with one expansion need have no relation to those chosen in connection with another. Thus one deals with the constraints imposed by each expansion separately.[187]

So far, we have considered only QC, but of course we are really interested in the consistency of arithmetical theories. What does (**Herb**) have to tell us about this?

Let L(A) be the language of a first order arithmetical theory, and let N be the standard model of arithmetic. With F now some sentence of L(A), and F^- as defined above, a *Herbrand evaluation for F in L(A)* is any structure S obtained from N by adding interpretations for the indicial functions of F^- and for the initial constant c. If $S \models E(F, p)$ for some p, then S is a *p-approximation* for F. The *p-range* of S is the set of numerical values t_S assigned by S to terms t in D(F, p). So: S is a p-approximation for F iff $S \models F^-$ whenever the variables in F^- are given values in the p-range of S.

We are now apparently speaking of interpretations and of the standard model of arithmetic, but (as Goldfarb points out) we have not invoked the full model-theoretic notion of truth in a structure. For a Herbrand evaluation involves no variables. Rather, we are invoking a *finitary approximation* for the notion of truth in a structure, which appeals to truth only with respect to the quantifier-free part of L(A). As a consequence of this, and of the way

---

[186] Plausible as it might be, the proof of this is not routine, and Herbrand's own treatment is deficient. See Dreben, Andrews, and Anderaa [1963].
[187] Goldfarb [1990] p53.

p-approximations are chosen, Herbrand evaluations always agree with standard evaluations in N of quantifier free formulas of L(A). Therefore, $S \models G(t_1, \ldots, t_n)$ iff $N \models (G(t_{S1}, \ldots, t_{Sn})$ for all quantifier-free formulas $G(x_1, \ldots, x_n)$, all terms $t_1, \ldots, t_n$ in an Herbrand domain D(F, p), and all Herbrand evaluations for F. The finitary approximation to the full model theoretic notion of truth in a structure really is a finitary approximation to *that* notion, since it is guaranteed to agree with it for quantifier-free formulas.

(Herb) now tells us that, in a given derivation from F, some p-approximation can simulate the use of quantifiers in that derivation, in a way which preserves truth of all quantifier-free formulas. Such a p-approximation assures us that it is possible to provide a finitary account of the workings of the quantifiers in that particular derivation, which involves no appeal to the full power of classical quantification. Herbrand's line of argument, then, keeps faith very closely with the central Hilbertian intuition that concentration on the structure of particular proofs can be made to yield a guarantee that any finitary formula proved with the use of 'ideal elements' - quantifiers - is 'finitarily true', or rather (and less misleadingly) recognizable as true by finitary means alone (in principle).

Unfortunately, this central Hilbertian insight can now be shown to be unconfirmable by finitary means, *precisely because it is true.* The demonstration that this is so will reveal an incompleteness in any theory T that is Σ_1-sound. To show this, Goldfarb states, and sketches the proofs of, the following series of results.

Σ_1-Lemma: if $(\exists x_1) \ldots (\exists x_n)G(x)$ is any Σ_1-sentence derivable from F, then there exists p (computable from the derivation) such that any p-approximation for F contains in its p-range a numerical value for x that makes G(x) true.

Proof sketch (Goldfarb): since the theories we are interested in will contain a function symbol for the p.r. pairing function, we may restrict our attention to sentences $(\exists x)G(x)$ with a single quantifier without loss of generality. Assume then that $QC \cup \{F\} \vdash (\exists x)G(x)$. Then by **(Herb)**, $E(F \wedge \neg(\exists x)G(x), p)$ is truth functionally unsatisfiable for some value of p. The functional form of $\neg(\exists x)G(x)$ is $\neg G(x)$. $\neg G(x)$ contains no function signs other than those occurring in the Herbrand domains of F, whence $D(F, p) = D(F \wedge \neg(\exists x)G(x), p)$.

But then the Herbrand expansion $E(F \wedge \neg(\exists x)G(x), p)$ is truth functionally equivalent to $E(F, p) \wedge \bigwedge \neg G(t)$, where the conjunction $\bigwedge \neg G(t)$ is $\neg G(t_1) \wedge \ldots \wedge \neg G(t_n)$ for all t_i

in D(F, p). By (**Herb**), then, E(F, p) $\wedge \bigwedge \neg G(t)$ is truth functionally unsatisfiable. Then, if S is a p-approximation for F, we have $S \models E(F, p)$, and so $S \models G(t)$ for some term in D(F, p). Therefore $N \models G(t_S)$. □

As a consequence of the Σ_1-Lemma, if there exist p-approximations for F for all p, then every Σ_1 sentence derivable from F is true in the standard model of arithmetic. The reason extends immediately to Π_2 sentences - sentences $(\forall x)(\exists y)G(x, y)$ with G(x, y) quantifier free. For if $QC \cup \{F\} \vdash (\forall x)(\exists y)G(x, y)$ with G(x, y) quantifier free, then $QC \cup \{F\} \vdash (\exists y)G(c, y)$ where c is the initial constant included in D(F, p) for each p. But then $(\exists y)G(c, y)$ is a Σ_1-sentence in which no function constants occur which do not occur in the Herbrand domains of F. Then by the Σ_1-Lemma, there exists some p for which any p-approximation for F has in its p-range an integer n such that $N \models G(c_S, n)$. This proves the

Π_2-Lemma: if $QC \cup \{F\} \vdash (\forall x)(\exists y)G(x, y)$, then there exists a p, computable from the derivation, such that, for any q, each p-approximation for F with initial value q contains in its p-range an integer n with $N \models G(c, n)$. □

Consequently, a procedure which effectively generated a p-approximation for given axiom(s) F with initial value q, for any p and any given q, would give an effective procedure for computing a choice function for any Π_2-sentence derivable from F. What Goldfarb (following Kripke) then shows is that the natural formalization of the sentence: 'for any p, q, there exists a p-approximation for F with initial value q' cannot itself be derivable from a consistent set of axioms F. For the relation {<p, q>: there exists a p-approximation for F with initial value q} is computable, therefore recursive by Church's Thesis. Then there is a formula G(p, q, n) of L(A) such that there is a p-approximation for F with initial value q iff $N \models (\exists x)G(p, q, x)$, where the correct value of x encodes the required p-approximation for F with initial value q, *whose p-ranges contain only integers less than x*. The sentence $(\forall p)(\forall q)(\exists x)G(p, q, x)$ then says: for every value of p, q, there exists an x such that x codes a p-approximation for F with initial value q.

The *diagonal sentence* $(\forall p)(\exists x)G(p, p, x)$ then says: for each p, there exists an x such that x codes a p-approximation for F with initial value p. Then

> **(Gold 1)** if for every p there is a p-approximation for F with initial value p, then $(\forall p)(\exists x)G(p, p, x)$ is not derivable from F.

Proof (Goldfarb): Suppose it is derivable. Then by the Π_2-Lemma there exists a p such that every p-approximation for F with initial value p will have in its p-range an integer n such that $N \models (\exists x)G(p, q, n)$. Pick for S a p-approximation for F with initial value p such that the largest integer in the p-range of S is as small as possible. Then there is an i in the p-range of S such that $N \models (\exists x)G(p, q, i)$. But by the choice of G(p, p, x) described above, there is a p-approximation for F with initial value p whose p range contains only integers less than i. Contradiction. □

The extension to infinite axiom systems proceeds as one would expect. Let T be an infinite axiom system, and for each n, let Tn the conjunction of the first n axioms. Then, given a p.r. listing of the axioms, the relation {<n, p, q>: there exists a p-approximation for Tn with initial value q} is computable, therefore recursive. Now argue as before, with $(\forall x)(\forall p)(\exists z)G(x, p, p, z)$ for the underivable sentence. The result confirms

> **(Gold 2)**: if, for every p, n, there exists a p-approximation for Tn with initial value p, then $(\forall x)(\forall p)(\exists z)G(x, p, p, z)$ is not derivable in T.

Goldfarb then goes on to use these facts to prove two more theorems which bear a more obvious resemblance to the familiar First Incompleteness Theorem of Gödel, but I think that the point has been sufficiently made. Commenting on **(Gold 1)** and **(Gold 2)**, Goldfarb writes

> The Π_2-sentences we have constructed are underivable because their choice functions outstrip, for each p, the power of p-approximations. The measure of this outstripping is an obvious one: the maximum of the integers in the p-range of the approximations. A choice function for a derivable Π_2-sentence can be obtained from p-approximations for fixed p and varying initial values. However, since our sentences assert the existence of p-approximations for each p, their choice functions simply grow too fast.[188]

There is, then, a measure of irony in the situation. Hilbert's central insight, that any derivation can only draw upon a finitistic amount of the full infinitistic power of the quantifiers occurring in that derivation, has been vindicated, but the vindication demonstrates that that very insight cannot itself be finitistically verified. At each stage, we

---

[188] Goldfarb [op cit] p58-59.

can state with great precision what is required for verification, but what is required for verification always just outstrips the reach of finitary approximations.

This way of arguing that the First Incompleteness Theorem shows that Hilbert's Programme cannot be carried out seems to me far preferable to some alternative suggestions that have been prominent in the literature, in that it makes no appeal to the content of the undecidable sentences. We may usefully take some time to consider a couple of these alternative suggestions.

In his indispensable survey of the incompleteness theorems Smorynski [1977], Smorynski claims (correctly) that (**Gödel 1**) shows that Hilbert's *Conservation Programme* must fail. He writes as follows:

> Hilbert's Programme can be described thus: There are two systems . . . F and I of mathematics. F consists of the finite, meaningful statements and methods of proof and I the transfinite, idealized such statements and methods. The goal is to show that, for any meaningful assertions P, if P is provable in I, then P is provable in F. . . . Gödel destroyed Hilbert's Programme with his First Incompleteness Theorem by which he produced a sentence G 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.[189, 190]

A striking feature of this criticism, though, is the claim that the sentences shown to be undecidable in extensions of **R** by Gödel's (or Rosser's) methods are said to be *true*. That claim was not made in our report of the logical facts surrounding (**Gödel 1**). Gödel himself makes no such claim (in Gödel [1931], at any rate). And no appeal need be made to the content of the underivable sentences constructed in Goldfarb's argument. In a

---

189 Smorynski [1985], pp3-4.

190 It is to be understood here that Smorynski is giving an informal, intuitive gloss on an argument for which he then goes on to provide in much greater detail. I therefore think it a little ungenerous of Detlefsen, in his initial response to this passage, to make much of the fact that the claim made in this passage is not strictly and literally true (see the Appendix to Detlefsen [1986]). To produce a direct counterexample to the Conservation Programme, one must indeed display a finitary system **F**, an ideal extension **I** of **F** which we have good reason to believe to be consistent, and a 'real' sentence **G** in the common language of **F** and **I** which is unprovable in **F**, and provable in **I**; and it is perfectly true that the particular result proved in Gödel [1931] does not do this, since, in that paper, Gödel produces an instance of a sentence undecidable in what Hilbert would have classified as an *ideal* system - in effect, the ideal system (**PA**). But the phrase 'the first incompleteness theorem' does not now standardly refer to that particular result. Rather, it refers our (**Gödel 1**). And a standard proof of (**Gödel 1**), using (**DIAG**), can indeed by guaranteed to provide a formula which 'satisfies a sufficiently narrow criterion of meaningfulness', which is not provable in **PRA**, and which is provable in consistent 'ideal' extensions of (**PRA**).

different place, Smorynski gives his reasons for claiming that '17 Gen r' and its correlates are indeed true as follows:

> [. . . the first incompleteness theorem destroys the Consistency Programme] since (1) the statement φ [i.e. '17 Gen r', or some other undecidable formula] is real; and (2) φ is easily seen to be true. ((1) requires looking at the construction of φ; (2) is seen by observing that φ asserts its unprovability and is indeed unprovable.) Thus, the First Theorem shows that the Conservation Program cannot be carried out . . ..[191]

This way of arguing from the First Incompleteness Theorem to the failure of Hilbert's Programme has sparked a rather bemusing series of debates.

Resnick, for example, denies that the problematic undecidable sentences are 'real', i.e. are of finitary character.[192] If we take the sentence used in (GÖD) as our specimen undecidable sentence, i.e.

$$(\forall y)\neg(\mu(y, \ulcorner G \urcorner))$$

then, according to Resnick, this sentence is not real. Rather, the *schema*

$$\text{(a)}\quad \neg(\mu(\mathbf{n}, \ulcorner G \urcorner))$$

(with **n** a schematic numeral) is real, and is regarded by the finitist as a 'metalinguistic device for communicating indefinitely many real sentences in one breath'.[193] Now, for each number n, we do in fact have

$$\mathbf{PRA} \vdash \neg\mu(\mathbf{n}, \ulcorner G \urcorner).$$

In virtue of the ω-incompleteness of **PRA**, though, we do not have

$$\mathbf{PRA} \vdash (\forall y)\neg(\mu(y, \ulcorner G \urcorner)).$$

unless **PRA** is inconsistent.

---

191 Smorynski [1977] p825.
192 Resnik [1974]pp119-121.
193 Resnik [op cit] p117.

But the same point is perhaps more revealingly put as follows. Whilst we can derive every *numerical instance* of (a), we cannot have

$$\mathbf{PRA} \vdash \neg(\mu(y, \ulcorner G \urcorner))$$

with *free variable* y, unless **PRA** is inconsistent. As you might say, one can prove every instance of the schema, but not the schema itself. The quantifier, in this metalinguistic way of putting the point, cannot pass through the turnstile into the universal quantifier of **PRA**.

The problem is, though, that this way of conducting the discussion is now apt to bog down in some nice distinctions between free variables and schematic letters, and this seems to me quite unsatisfactory. The important point, surely, is that the underivability of the last formula displayed above shows that there is something amiss with Hilbert's notion of generality. The classical quantifier does not in fact have the character Hilbert hoped. Classical mathematics, with the classical quantifier, is not conservative over **PRA**, and this is clearly shown in the independence of $(\forall y)\neg(\mu(y, \ulcorner G \urcorner))$ in **PRA**. It seems to me that Goldfarb's argument makes this point clearly, without this detour through disputes about free variables, schematic letters, and real formulas.

In an equally unsatisfactory fashion, Detlefsen has engaged Smorynski in a dispute over whether the notion of *truth* involved in Smorynski's claim that '17 Gen r' and its correlates is the 'classical' notion of truth, or some (unspecified and unexplained) 'finitary' notion of truth. Now, I should have thought that there is no need to introduce a notion of finitary truth - or a special notion of classical truth, for that matter - in order to assess the feasibility of what Hilbert was trying to do (no such notion is to be found in Hilbert, for one thing). In my flatfooted way, I should have thought that a finitary truth is just a truth - an ordinary truth - of finitary mathematics. There is no need for more than one notion of truth in the discussion: we are simply required to respect a distinction between truths which can and which can not be recognized with the very restricted resources of finitary mathematics. But however one feels about this notion of finitary truth, it is not needed in order to show that Hilbert's Programme cannot be carried out with full generality. For the Kripke-Goldfarb argument, sketched above, establishes that, without any appeal to the content of the undecidable sentences.

And this is as it should be. The First Incompleteness Theorem is a purely syntactic result, and the syntactic facts alone suffice to show that the Master Argument cannot be carried out

with full generality. No finitarily verifiable procedure can be given for the elimination of ideal elements - *classical* quantifiers - from proofs of finitary formulas. This is not to deny, notice, that this part of Hilbert's Programme can in fact be carried out to a surprising degree - recall the brief discussion of Friedman and Simpson's work in **Chapter One**. But the full generality to which Hilbert aspired is not to be had.

We might put the fundamental point like this. There are in fact *two aspects* of the proof of the incompleteness of elementary arithmetic. First, there is the definability of Gödel's recursive substitution function (**Subst**), satisfying

$$\textbf{(Subst)}:\ A(\ulcorner F(x) \urcorner, n) = \ulcorner F(n) \urcorner$$

(i.e. (subst) gives us as value the code number of the result of substituting **n** for **x** in F(x)). With **(Subst)** in hand, we prove the Diagonal Lemma as follows.[194] Let F(x) be any formula with x alone free. Let B(y) be F(A(y, y)), with the code number m. Finally, let G be B(**m**). Then with **T** as above,

$\mathbf{T} \vdash G \leftrightarrow$	$F(A(\mathbf{m}, \mathbf{m}))$	(since G is $F(A(\mathbf{m}, \mathbf{m}))$)
$\leftrightarrow$	$F(A(\ulcorner B(y) \urcorner, \mathbf{m}))$	(since m is the code number of B(y))
$\leftrightarrow$	$F(\ulcorner B(\mathbf{m}) \urcorner)$	(since $A(\ulcorner B(y) \urcorner, \mathbf{m})$ is $\ulcorner B(\mathbf{m}) \urcorner$ by **(subst)**)
$\leftrightarrow$	$F(\ulcorner G \urcorner)$	(since G is B(**m**))

This aspect of Gödel's procedure in Gödel [1931] makes the argument of that paper appear to be very close to the semantic paradoxes, for the use of **(Subst)** to produce an underivable formula that 'says of itself that it is not provable' is made to appear crucial to the proof.[195] Actually, Rosser's refinement of Gödel's result is already enough to make one wonder about this, since the underivable sentence **(ROSS)** certainly does not 'say of itself' anything at all. But the Kripke-Goldfarb argument sketched above sharpens this doubt.

---

[194] The following sketch of an argument is taken from Smorynski [1985], p6.

[195] The highly misleading impression that the central argument of Gödel [1931] is closely related to the Liar paradox is of course reinforced by Gödel's informal introductory discussion - a discussion which, I suspect, Gödel may have come to regret. One sometimes gets the impression, in philosophical discussions of Gödel's work, that this is the only part of the paper that has been read.

For that argument concentrates entirely on the other, genuinely profound aspect of Gödel's proof - the process of Gödelization, of assigning codes to syntactic expressions, which enables us to pull the syntax of a given mathematical theory down into the theory itself in such a way as to associate derivable formulas with syntactic facts. The Kripke-Goldfarb argument shows that *this aspect of Gödel's work is sufficient for incompleteness*, and in particular sufficient to show that the finitary elimination procedures Hilbert hoped for are not to be had. In Kripke-Goldfarb, there is no use of a substitution function to manufacture a 'self-referential' sentence. There is indeed a crucial use of diagonalization - of the identification of two arguments of a relation - but not of self reference. And this has the added attraction of setting aside the misleading suggestion of a close resemblance between incompleteness and the semantic paradoxes suggested by Smorynski's anti-Hilbert argument.

In stressing the purely syntactic nature of the difficulties for Hilbert's Programme revealed by the First Incompleteness Theorem, let me emphasize that nothing I have said is intended to dispute the existence of true, 'real' sentences that are not provable in **PRA**. The above remarks are not intended to imply that I have any serious reservations about the familiar claim that the sentence '17 Gen r' of Gödel's original proof is true 'because it says that it is unprovable, and so it is'. (It happens that that claim is strictly and literally false, but that is a quibble.) I am simply trying to emphasize that the impossibility of fully attaining Hilbert's goals can be shown without engaging in these kinds of semantic issue.

On the kind of Hilbertian position I outlined in **Chapter One** and **Chapter Two**, the truth predicate is to be treated disquotationally. Now, I think that **PRA** is consistent, so I think that it is true that **PRA** is consistent (or, better, I think that '**PRA** is consistent' is a true sentence of English). I also think - as did Hilbert - that the question of the consistency of **PRA** is finitarily meaningful. I therefore think that '$\mathrm{Con}_{\mathbf{PRA}}$', where that sentence is constructed from a provability predicate for **PRA** in the standard way, is true, and is not provable in **PRA**. I therefore think that there are true sentences of finitary character which are provable in infinitary extensions of **PRA**, yet not provable in **PRA** itself. I hold, therefore, that the Conservation Programme fails, definitively. I also think that the impossibility of finitarily establishing the existence of finitary surrogates for the classical quantifiers in proofs of finitary sentences, as demonstrated in the Kripke-Goldfarb argument, gives the fundamental reason why Hilbert's Programme cannot be carried out in full generality.

But nothing that I have conceded here gives grounds for any revision of the Hilbertian position on our ontological commitments in mathematics outlined in the earlier parts of this essay. If a Hilbertian thinks that conceding the truth of, say, 'Con_{PRA}' will force upon her ontological doctrines of a kind uncongenial to Hilbert, then I think that she must be enmeshed in the conflation of issues concerning the acceptability of scientific theories and issues concerning ontological commitments we discussed in **Chapter One**. 'Con_{PRA}' is provable in a mathematical system we have very good reason to accept, as well as being intuitively obvious. Coupled with a good philosophical account of the truth predicate, and a good theory of belief, this will enable anyone to assert with a clear conscience: 'Con_{PRA}' is true, putting her mind where her mouth is. The thought that such an admission will *of itself* force us into 'ontological commitment' to, say, the iterative hierarchy of sets seems to me to be plainly misguided. What is required in order to force such an admission is a heavy investment in controversial positions in the philosophy of language and the philosophy of mind. And here, I think, the Hilbertian can view her prospects of defending her corner with some optimism.

Section Five: Hilbertian Philosophy of Mathematics and Incompleteness. I have now accepted that the incompleteness theorems show that Hilbert's Programme cannot be carried out with full generality, and given my reasons for believing this. In somewhat misleading brevity, the reason is that Hilbert was deeply wrong about the nature of the (classical) quantifier, and as a consequence of this, we should conclude that the infinite of classical mathematics does not have the character that Hilbert thought it had. This suggests to me that, in some sense, Hilbert's Programme founders on the nature of classical logic, rather than classical mathematics. Be that as it may, founder it does, and we should now consider further the implications this has for the modified Hilbertian philosophy of mathematics I sketched in chapters **One** and **Two**.

Occasionally, one sees suggestions that (**Gödel 1**) undermines Hilbert's Programme by revealing a fundamental shortcoming of the axiomatic method. This is explicitly argued in a recent book by A.W. Moore. He writes:

> Why is [the project to axiomatize mathematical theories] threatened [by (**Gödel 1**)]? Because what Hilbert had envisaged - at least as a paradigm - was a single, complete axiomatization. And Gödel's theorem shows that nothing matches that paradigm. Any axiomatic base for transfinite mathematics must needs be supplemented. Not only that, but there will be one particular way of supplementing it that seems forced upon us; and this casts doubt on the idea that only finitary propositions genuinely describe mathematical reality. What will seem to force us to supplement the base in one way rather than another will be

> non-finitary reflection on the consistency of the base - reflection on the fact that, in the infinite landscape with in which the base is located, there are no paths leading from it to each of two contradictory statements.[196]

Now, I know of no evidence for the assertion that Hilbert envisaged (as a paradigm) a single, complete axiomatization. I suspect that this is a confused echo of a claim Hilbert really did make, that all of classical analysis can be formalized in Z_2. *And that claim is true*. But we shall let exegetical matters pass, along with Moore's tendentious talk of reflecting on infinite landscapes. For there surely is a genuine worry lurking behind this passage. After all, one of the things that we know as a consequence of (Gödel 1) is that arithmetic is not axiomatizable. Does that not suggest a fundamental limitation of the axiomatic method?

On reflection, I think one ought to answer, No, it does not. For although the import of the first incompleteness theorem is indeed that the notions of *truth in the standard model of arithmetic* and *derivability in PA* are not co-extensive, our understanding of what that in fact ***means***, and our grounds for believing it, are given to us very largely by our understanding of a further mathematical theory with respect to which, more than any other, the merits of the axiomatic method have been forcibly demonstrated. I mean, of course, set theory, and in particular **ZF**. But perhaps we should go over this ground a little more slowly.

To begin with, let us observe that, as a description of mathematics as it is actually practised, the central claim being made by Moore in the above passage is simply false. Almost all of contemporary mathematics - almost every area within the sixty or so major divisions and thirty four thousand subdivisions of contemporary mathematical enquiry - can be and standardly is formalized in ordinary **ZF**.[197] What is more, we have already seen that the ordinary mathematics that Hilbert most cared about can be formalized in systems that are far, far weaker than full **ZF**. So far from feeling the need to 'supplement' the axiomatic base provided by **ZF**, in order to get an adequate scientific grasp of ordinary mathematics, a great deal of mathematical research is aimed at finding *weaker* systems adequate for standard mathematical practice. We saw in **Chapter One** that research of this kind has met with some spectacular successes.

---

[196] Moore, A.W. [1990] p178.

[197] The statistics are based on the 1979 classification of mathematics of the *Mathematical Reviews* - see Davis, P.J, and Hersh, R. [1980] pp29-30. The subject has, of course, grown in the last nineteen years, without falsifying the claim that almost all of mathematics can be formalized in **ZF**.

But what of the claim that we are forced to 'supplement' any given axiomatic basis for mathematics by the demands of reflection on the consistency of that basis? Again, over the vast bulk of mathematics as it is actually practised, this claim is once again simply false. Issues of consistency within mathematics are standardly regarded as resolved once it has been shown that the axiom system in question has a model - where this standardly means, has a model in set theory. The procedures employed here are refinements of those pioneered by Hilbert in his early studies of geometry and analysis. I know of no evidence that they are regarded by mathematicians as in some way intrinsically limited by the incompleteness theorems, nor can I see anything in the incompleteness theorems that ought to make one so regard them. Consistency, together with such closely related issues such as independence, completeness, categoricity etc., fall within the province of model theory. And model theory is an axiomatic discipline - indeed, it is an axiomatic discipline not clearly to be distinguished from set theory.

But what about 'reflection' on the consistency of set theory itself? Moore writes:

> It is all very well drawing 'V-shaped diagrams intended to capture our intuitions about what Sets are like. But our most basic intuitions in this area have already proved unreliable. . . . Is there any way of guaranteeing ZF's consistency without relying on our intuitions about what Sets are like? It seems not, given Gödel's theorem. The lesson here is of a piece with the lesson that faced the finitist. If we are going to talk about the infinite in a mathematically precise way, then we really must see ourselves as *talking about the infinite*. And if we are going to ratify what we say, then we can make do with nothing less than insight into what the infinite is actually like.[198]

But once again, most of what is claimed here is simply false. 'Our' most basic intuitions about sets have not been proved to be unreliable. The so-called 'logical' conception of set implicated in the paradoxes was introduced into foundational studies by a *philosopher*, engaged upon a philosophical project, and was in fact regarded with suspicion by mathematicians almost from the inception of naive set theory. Frege himself was uneasy about the fatal Axiom V of the *Grundgesetze*, and we saw in **Chapter One** that the Göttingen mathematicians were already well aware that there was something badly wrong with the naive abstraction principle by the later nineteen nineties.[199] So far from casting doubts upon the 'intuitions' of mathematicians, the early history of set theory provides a

---

[198] Moore [op cit] p179.

[199] For Frege's 'intuitions' about Axiom V, see Grundgesetze vol II, Appendix, p253 in the German edition (English translation in P. Geach and M. Black (eds) p214).

major vindication of them. Despite the suspicions surrounding abstraction, almost all the mathematical work done in naive set theory before Zermelo's axiomatization remains acceptable today with minimal modifications. Hausdorff's 1914 text is still in print and still regularly consulted by mathematicians, despite the fact that it is written in the naive, 'intuitive' tradition of set theory purportedly shown to be 'unreliable' by the paradoxes. Virtually nothing of consequence in Cantor's own work has had to be abandoned - for the simple reason that the 'intuitive' understanding of set theory that really had been guiding the practice of the great majority of mathematicians (including Cantor) from the beginning - the conception of sets as the product of an iterative process of generation - has never given rise to grounded suspicions of inconsistency.

What is more, mathematical reflection on the *consistency* of set theory has not had the character that Moore suggests in these passages. 'Intuition' about what the infinite is 'really like', whatever that might mean, has had little or no role to play. Our present high degree of confidence that **ZF** is consistent is in fact due to a series of studies that show the consistency of set theories which include those set-theoretic axioms about which we feel some degree of intuitive uncertainty - most notably, the axiom of choice - relative to fragments of set theory that do not include those axioms. The paradigm here is the proof that the axiom of choice is independent of the rest of **ZF**: if **ZF** minus choice is consistent, then **ZF** is consistent. The model theoretic investigations that have established this result have not proceeded by *supplementing* **ZF** on the basis of some alleged 'insight' into the transfinite. Rather the opposite: they have typically proceeded by constructing so-called 'inner models' of the **ZF** axioms, models in structures *simpler* (and 'smaller') than the structure of **ZF** itself.

There is also something very peculiar about the implied dichotomy between reliance on 'intuition' and reliance on axiomatization. To begin with, nothing in the axiomatic method, as Hilbert understood it, is incompatible with the evident need for some intuitive grasp (in a philosophically uncontroversial sense) of the content of a mathematical theory, real or ideal. We do indeed have some intuitive grasp of set theory, and this intuitive grasp is expressed in the axioms of **ZF**. We may well have, or we may well develop some intuitive grasp of the mathematical import of extensions of **ZF**, and we express the import of that intuitive grasp by the addition of large cardinal axioms or whatever. Hilbert's advocacy of the axiomatic method was not intended to establish that mathematics had no need of a notion of 'insight'. Hilbert simply believed that the progress of science - of *science as a whole*, not just mathematics - would be impeded unless it was guaranteed that all the 'insight' that the

practice of science required could be regarded as *internal* to scientific practice, rather than being imposed upon science from without on the basis of some philosophical doctrine - as in Frege, say, or Kronecker. Axiomatization is intended only to ensure that appeals to 'intuition' are kept within intersubjectively manageable bounds - so as to avoid the impasse that had so long impeded the study of geometry. It is for this reason that the Hilbertian insists that any intuitive understanding of a mathematical notion is to be expressed in axiomatic form, and ratified by metamathematical and other *mathematical* studies of the mathematical properties of the resulting axiomatic system.

Now, since any theory in natural science will surely have to include enough arithmetic to make the representation of the primitive recursive functions possible, we may assume that **(DIAG)** can be proved with respect to any natural scientific theory. Could anyone seriously suggest that the inadequacy of the axiomatic method in physics had been shown by Gödel's proof that any axiomatized *physical* theory must be incomplete? Are we ever treated to disquisitions on the mysteriousness of our grasp of physical reality on the grounds that we can 'intuitively' see that some physical sentences are true even though they are not derivable in physical theory? These suggestions are frivolous. The adequacy of a physical theory is to be judged by the physicists on the basis of the standards - of empirical adequacy, simplicity, fruitfulness etc. - commonly employed for that purpose, and those standards simply cannot be brought to bear unless we can see how to formalize the theory in question axiomatically. The fact that an entirely general mathematical argument can be then be deployed to produce sentences true in the intended physical model of the axiomatic physical theory but underivable from the axioms would not be regarded as indicating a serious shortcoming in the axiomatic method in physics, and I can see no reason to think that the practice of the physicists is in any way misguided in this respect. I am equally unable to see that the fact that we have a *mathematical guarantee* that any interesting mathematical theory is incomplete poses any problem whatsoever for the accepted practice of mathematicians, which is equally dependent upon axiomatization before the standards of acceptability of mathematical theories can be deployed.

However, it might seem that this response fails to get at the roots of Moore's worry. For does not **(Gödel 1)** appear to establish the existence of an unbridgeable gap between the notions of *truth* and *derivability* in mathematics? Have we not been shown that any attempt to display the truths of mathematics in an axiomatic system or series of axiomatic systems must necessarily fail?

Now, I think that there is indeed a genuine worry in this area, but it is not well expressed as a worry about the axiomatic method. For in general, the assignation of a semantic value to sentences produced by (DIAG) is via a model-theoretic argument, and model theory, I say again, is no less of an axiomatic discipline than the rest of mathematics. Indeed, the import of (DIAG), not only for model theory but for semantic theory in general, is surely that semantic theory *must* proceed axiomatically, since the principle conclusion to which one is forced by (DIAG) is precisely that our naive semantic intuitions are not to be trusted. (DIAG) can be deployed in such a way as to derive contradictions from the natural principles governing each of the central semantic notions - truth, satisfaction, and reference.[200] Given that this is so, it is really a very bad idea to use some 'intuitive' notion of the correct semantic evaluation of sentences produced by diagonalization as a stick with which to beat the axiomatic method. It is a still worse idea to leave the acceptability of any mathematical results hostage to some kind of 'insight' independent of the practice of mathematics, in the manner that Moore suggests. The evident consequence of adopting such a position will be the return to the kind of futile, scientifically sterile deadlock one finds in Frege's rejection of non-Euclidean geometries, or indeed in the pre-war perplexities over the 'interpretation' of quantum mechanics.

Nevertheless, the incompleteness phenomena do show that there is something deeply wrong with the conception upon which the sole ground for assertibility in mathematics is that provided by derivability from 'arbitrarily postulated' axioms. But that was never Hilbert's view, nor is it the view of the Hilbertian philosophy of mathematics sketched in the earlier parts of this essay. For my Hilbertian, our understanding of finitary mathematics is not simply a matter of our grasp of the deductive consequences of arbitrarily postulated axioms. Rather, mathematical concepts at this level are integral to a kind of fragmentary and naively given physical theory which is a presupposition of any theory of representational thought. Acceptability at this level is not simply internal to the practice of mathematics, since mathematics at this level - as Hilbert himself insisted - is not clearly to be distinguished from the neighboring disciplines of physics and what Hilbert called 'epistemology'. The grounding for our understanding of mathematical concepts provided by this basic level weakens as the seas of mathematical complexity mount, and it is a feature of my Hilbertian's position that, at the higher levels of complexity, the meaningfulness of mathematical notions becomes almost entirely internal to the practice of mathematics. Here, the picture of mathematical meaning as grounded solely in the

---

200 See McGee [1991] Chapter 1 for details.

deductive consequences of 'arbitrarily postulated axioms' is much more plausible, nor is it under any threat from the incompleteness theorems. For there are reasons independent of the incompleteness theorems for rejecting the idea that research at the outer boundaries of set theory is best thought of as the search for truths potentially independent of our mathematical capacities. We mentioned some of those reasons in the discussion of Maddy's views.

Of course, there remains the point that, contrary to Hilbert's hopes, the consistency of almost all interesting mathematical theories cannot be finitistically verified. I have downplayed Hilbert's interest in consistency for its own sake in this essay, for I think that the Conservation Programme is both the most important and interesting aspect of the historical Hilbert's foundational thought, and the heart of Hilbert's Programme as I understand it. Nevertheless, Hilbert did hope for a finitary proof of consistency, and I have accepted the common view according to which Gödel has left us with no such hope. **PRA** *is* consistent, and it seems to me that its consistency is transparent in, say, Goodstein's formalization of PRA as a 'logic free' equation calculus. Still, the proof theoretic verification of the consistency of PRA, thus formalized, is not finitary in character. This is, I think, an astonishing fact, all the more so if one has been attracted by the Hilbertian intuition that inspired the substitution method. Denied a finitary consistency proof for finitary mathematics, my Hilbertian at least has the consolation that a great deal of mathematics, including a great deal of infinitary mathematics, can be got within the confines of systems that are conservative over **PRA**. Given that at least part of Hilbert's desire for a purely finitary consistency proof had its origins in some rather muddled views on the ontological commitments of mathematics, that consolation is, I think, far from negligible.

APPENDIX ONE:
Hilbert and the Philosophy of Mathematics

Hilbert was a mathematician, not a philosopher. In this, he contrasts sharply with Frege and Russell, his great contemporaries in mathematical logic. Outside of pure mathematics, his main professional interest was in theoretical physics, a field to which he devoted much energy. To be sure, he also had an interest in philosophy, and he appears to have had something more than a casual acquaintance with the work of Kant. This much is perhaps to be expected of a German scholar of encyclopaedic scientific curiosity and outstanding intellect, in an age in which the fragmentation of the academy was not yet pronounced. But he had no training in philosophy, he had little contact with the philosophers amongst his contemporaries, and I know of no evidence that he ever made a systematic study of any philosopher or philosophical issue other than those forced upon him by his work in the foundations of mathematics.[201] And even there, Hilbert's attention was held only by problems he felt he could tackle by mathematical techniques. This is the first thing that has to be born in mind when reading Hilbert's more prosy writings.

However, those writings are often both indisputably philosophical in character and intent, and extraordinarily suggestive. In particular, it is clear that Hilbert's conception of finitary mathematics rests upon striking insights into the importance, not only for the foundations of mathematics but for our most general theory of cognition, of the recursive functions. In the central parts of this essay, I have attempted to describe a position in which those insights are accommodated, which is both reasonably clear and roughly in accord with Hilbert's intentions.

In any such attempt, there is surely a danger that Hilbert's writings become a mere vehicle for views which would be more appropriately presented as the authors' own. This danger is always present when the philosophers' eye is caught by the philosophically suggestive work of a scientist, and I have not even tried to avoid it. However, I have made efforts not to disguise, but rather to signal, those many places where Hilbert's own words are left far behind. Let it be clear, though, that I have made no attempt systematically to explore the

---

[201] The main source of biographical information on Hilbert is Reid [1986], but Blumenthal [1935] and Bernays [1967] are also valuable.

many ways there are of developing Hilbert's compressed and fragmentary philosophical ideas. This essay contains just one way - an interesting and plausible way, in my view - of doing this.

Hilbert's philosophical writings are the consequence of his mathematical interests having drawn him towards a branch of mathematics - mathematical logic, then in its infancy - which was at that time not clearly to be distinguished from the best extant philosophical reflection on the nature of mathematics. This is no longer true. Mathematical logic is now a flourishing branch of mathematics, with problems formulable in a purely mathematical vocabulary and resolvable by purely mathematical techniques. That this is so is due, in no small measure, to the achievements of Hilbert and his assistants. Proof theory in particular has many thriving research programs, most if not all of them at a considerable distance from Hilbert's own foundational concerns.

For Hilbert himself, though, mathematical logic had a quite different aspect. The vocabulary and techniques with which we are now familiar had in large part still to be developed. What is more, the pressure to develop those techniques was provided by problems, such as those posed by the paradoxes of naive set theory, which were not yet perceived as purely mathematical problems, and which offered no very obvious purchase for the mathematicians' professional skills. A further thing to be born in mind by the philosophical reader of Hilbert's prose writings, then, is the fact that they form part of an attempt to advance foundational studies to a stage at which the properly mathematical content of foundational problems would be clearly visible, and the mathematical techniques required for their resolution commonly accepted and understood. It is very important to bear in mind that Hilbert's aim was to *minimize* the role of philosophy in foundational studies.[202]

A final exegetical point that demands more detailed attention concerns Hilbert's relationship to Kant. Hilbert often mentions Kant with approval, and in his writings on foundations - especially in those places in which the notion of intuition plays a prominent role - he often lapses into the technical jargon of the *Critique of Pure Reason*. In part, of course, that is attributable to the currency of that jargon in the kinds of philosophical writing on mathematics, and especially geometry, with which Hilbert would have been familiar. But

---

[202] In brief: the philosophical reader of Hilbert must always bear in mind his admonition to his fellow workers in mathematical logic: *Wir sind Mathematiker!*

in addition to this, it is clear that Hilbert really does take himself to be affirming a Kantian position on the nature of mathematics. Consider for example the passage that plays such a large role in **Chapter Two** above:

> Kant already taught - and indeed it is part and parcel of his doctrine - that mathematics has at its disposal a content secured independently of all logic and hence can never be provided with a foundation by means of logic alone; that is why the efforts of Frege and Dedekind were bound to fail. Rather, as a condition for the use of logical inferences and the performance of logical operations, something must already be given to our faculty of representation [in der Vorstellung], certain extralogical concrete objects that are intuitively [anschaulich] present as immediate experience prior to all thought. If logical inference is to be reliable, it must be possible to survey these objects completely in all their parts, and the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects, as something that neither can be reduced to anything else nor requires reduction. This is the basic philosophical position that I consider requisite for mathematics and, in general, for all scientific thinking, understanding, and communication.[203]

The resonances of the critical philosophy here are brazen. If we are to intepret Hilbert aright, we need to know what to make of these resonances.

Now, there are certainly two very broadly Kantian themes being sounded here. The first is the claim that mathematics has a content which cannot be 'secured' by logic alone. This is clearly directed against the logicism of Dedekind and Frege: as we have seen, Hilbert thought it no coincidence that the systems of those two writers become inconsistent at exactly the point at which an attempt is made to derive the content of mathematics from purely logical principles. Apparently, Hilbert took this failure to speak for a 'Kantian' view of mathematics, but so far as this point is concerned, we can take that to be quite simply a view on which mathematics has a distinctive content, as logic does not. That is undoubtedly a Kantian view, in the sense that it is a view that Kant held. But it is by no means *distinctively* Kantian.

The second Kantian theme is considerably more murky. It concerns the reliability of what Hilbert calls 'contentual logical inference'. The claim Hilbert makes about this is that the validity of 'contentual logical inference' depends upon the *surveyability* of the objects with respect to which the inference is being made, where surveyability, we may take it, is a property of finite (and perhaps denumerable) collections, but not of (non-denumerably) infinite ones. Now, at some risk of travesty, we can perhaps describe a central aspect of

---

[203] Hilbert [1925] p376.

the critical philosophy to be a defence of the claim that the antinomies of traditional metaphysics are to be attributed to the employment of concepts outside their domain of legitimate employment. For Kant, that domain (so far as finite minds are concerned) is given by the limits of possible experience: any attempt to put to cognitive use concepts which can have no application within the limits of possible experience for such minds (such as the concept of God, or of the immortal soul), according to Kant, is bound to lead to inconsistency. It is clear that Hilbert thinks that some analogy holds between the paradoxes of naive set theory and the Kantian antinomies of pure reason - both are supposed to be attributable to the use of concepts outside the limits of their legitimate application; but what the analogy is supposed to be, and how the Kantian defence of that claim carries over to the set theoretic case, are matters about which Hilbert says nothing.

Thankfully, I do not think that we shall need to make good this deficit on his behalf, for at anything beyond the level of this kind of extremely vague and general affinity, the Kantian resonances in Hilbert, I find, give out rather quickly.

Let us focus for a moment on the notion of intuition, for I think it illustrates rather well the general point that needs to be made here. 'Intuition' ('anschauung') is certainly a key term in Kant. Kant does speak of objects being 'given in intuition', and he does think that mathematical objects in particular have deep connections with intuition. Nevertheless, it seems to me evident that Hilbert's notion of mathematical intuition, such as it is, differs in several crucial respects from that of Kant.

In the first place, mathematical intuition as Hilbert understands it is very much more restricted in scope than its Kantian counterpart. Even if we confine our attention to the mathematics known to Kant, only a small fraction of it counts as intuitable in Hilbert's sense. For Kant, on the other hand, intuition is implicated in *all* (non-trivial) mathematical knowledge. Kant holds, quite generally, that mathematical objects are 'constructed' in pure intuition, and that intuition plays an indispensable role in all mathematical cognition. Thus the real line, for example (along with the objects of Euclidean mathematics) counts as intuitable by Kantian standards. Thus intuition, in Kant, is indispensable to our knowledge of the topological properties of the real line. For Hilbert, on the other hand, the real line belongs to ideal mathematics, and a scientific understanding of its topology proceeds via an axiomatization the primary purpose of which is to render appeals to intuition redundant.

So far, this might seem to be no more that a disagreement over the *range* of intuitive mathematics. And there are indeed passages which suggest that Hilbert thought he was in agreement with Kant over the *nature* of mathematical intuition, differing only over its *extent*. (In the case of geometry, as we shall see, Hilbert sometimes seems to say that projective geometry has the kind of status Kant wrongly attributed to Euclidean geometry.) But even if Hilbert did think this, it seems clear to me that he was quite wrong.

To begin with, Hilbert never speaks of mathematical objects as being 'constructed' in pure intuition. Indeed, he makes no use at all of the (crucial) Kantian distinction between pure and empirical (or sensory) intuition. When Hilbert speaks of mathematical objects being 'given' to us 'prior to all thought', he apparently has in mind the 'numerals' we discuss in **Chapter Two** - arrays of strokes, for example, such as //// (the Hilbert numeral for the number four). But these count as intuitable for Hilbert in the sense that you can write down or otherwise reproduce *physical exemplars* of them which are *literally*, not figuratively, perceptible - perceptible by *outer*, not inner sense. In Kantian terms, these numerals are objects given in *empirical*, not pure intuition. This is very remote from anything Kant has in mind when he spoke of intuition with respect to mathematics. Even if it is permissible to think of Kantian construction in pure intuition in terms of the production of some kind of mental image, the image constructed cannot be regarded as an image of any kind of physical object. What you construct in pure intuition, on Kant's view, really is a genuine Euclidean triangle (as it might be), and you construct it out of genuine line segments. These cannot be thought of as copies, or images, of anything physical.

Again, there is no trace in Hilbert's writings of the Kantian thesis that space and time are formal features of human sensibility. This thesis provides the grounding for the special status Kant assigns to mathematical knowledge, for, as is well known, Kant thinks that (Euclidean) geometry is a kind of systematization of the content of the pure intuition of space, and arithmetic a systematization of the content of the pure intuition of time. The allegedly special status of Kant's favorite mathematical theories, then, is accommodated by turning those theories into 'structural' features of cognition, features necessarily present in any experience of an objective world. These doctrines are not to be thought of as inessential components of Kant's views of mathematics: on the contrary, they are the very essence of those views, and no account of mathematical intuition that fails to accommodate them has any claims on the authority of Kant.

But there is little trace of this in Hilbert. In particular, nothing in Hilbert's copious writings on arithmetic suggests that arithmetic is to be thought of as having some special relationship, or indeed any relationship at all, to the 'intuition' of time. One finds something like this in Brouwer, but this is an aspect of Brouwer's thought that Hilbert detested. In our discussion of the Hilbert/Frege correspondence in **Chapter One** we saw that Hilbert was very hostile to attempts to found mathematics on metaphysical doctrines of this kind.

As a consequence of all this, *nothing in Hilbert commits him to some position on the synthetic apriori character of mathematical judgments.* Yet the claim that mathematical judgments are typically synthetic apriori is absolutely central to Kant's views on mathematics, and Kant's claim that the objects of mathematics are constructed in pure intuition is above all else a defence of the synthetic a priori character of mathematical judgment. No such demand is to be made of Hilbert's notion of mathematical intuition. Whatever Hilbert means by 'intuition', then, it is not the 'pure intuition' of the First Critique.

The point made at length here with respect to the notion of intuition is in fact quite general. Although some key words from the critical philosophy crop up in Hilbert's writings, and although there is indeed some very vague and general affinity with Kantian thought to be found in Hilbert, the surrounding Kantian structure which gives the jargon of the critical philosophy its distinctive philosophical content is almost entirely absent. What is more, nothing that Hilbert actually says *depends* upon any part of the critical philosophy. Importing that surrounding structure, in my view, does little or nothing to illuminate or support any characteristically Hilbertian thesis, and some important Hilbertian theses, such as those concerning the nature and intuitability of finitary mathematical objects, seem to me flatly incompatible with Kant.

It is therefore a mistake, in my opinion, to treat Hilbert as in any serious sense a follower of Kant. To do so is to visit upon his writings a kind of general philosophical sophistication they simply do not have, and a host of philosophical liabilities they need not incur.

Getting a correct perspective on the relationship between Hilbert and Kant is an important step towards getting clear on the most fundamental, and the most complex questions concerning the motivations behind Hilbert's Programme. All too often Hilbert's

Programme has been thought to be simply an attempt to provide mathematics with an unshakeable foundation, proof even against the tremors emanating from the discovery of the paradoxes of naive set theory. When seen in this light, the importance of a proof of consistency for a mathematical theory seems to lie in its effectiveness against skepticism, and the importance of a finitary proof of consistency is thought to derive from the peculiar indubitability of finitary mathematics. Willingness to see Hilbert as a disciple of Kant fits this perspective nicely, for is not Kant the great champion of the indubitability of mathematics?

A recent paper by Philip Kitcher provides a paradigm of this approach, and it is not a coincidence that Kitcher attacks Hilbert's Programme with arguments very similar to those he uses in a companion paper on Kant's philosophy of mathematics.[204] According to Kitcher, Hilbert is best thought of as making a botched attempt at expounding the critical philosophy of mathematics. After announcing that the aim of Hilbert's foundational programme was to 'defend the thesis that we can have certain mathematical knowledge', Kitcher goes on to attribute Hilbert's alleged conviction that certainty was possible in mathematics to his acceptance of the following claim:

> We can obtain [the right to feel convinced about some mathematical statement] because we are able to give a special type of justification for mathematical claims. The hallmark of this type of justification is its absolute reliability. Once a mathematical claim has been justified in the special way, nothing can count as evidence against it . . ..[205]

Later, Kitcher makes clear his view that this 'special way' is in fact by intuition. He writes:

> According to Hilbert, *intuition and intuition alone* can yield the basis for certain mathematical knowledge.[206]

There are, however, many objections to this interpretation.

The least of them is simply this: Hilbert *never* claims that we can have 'certain' mathematical knowledge. Nowhere does he say that the aim of his foundational programme is to restore lost certainty to mathematics. To be sure, he claims, repeatedly, that mathematics has always been the paradigm of 'reliability and truth', and talks as if this

---

[204] See Kitcher [1975] and [1976].
[205] Kitcher [1976] p99.
[206] Kitcher [op cit] p106, my emphasis.

status was under threat from the paradoxes.[207] He writes, 'where else would reliability and truth be found if even mathematical thinking fails' (Hilbert [1925] p375). He is fond of talking of the 'security' of inferences in elementary number theory, and this 'security' is undoubtably to be attributed, in Hilbert's view, to the 'intuitive' character of that subject. But (as I shall argue in more detail below) 'intuitive' here just means, *obvious*. When he talks of the deliverances of mathematical intuition, he will indeed speak of 'immediate clarity', of 'reliability': but he does not speak of certainty. He does not speak of immunity to doubt in any philosophically committed sense. Nor does he ever suggest that there is any deep difference in the epistemological status of mathematical and physical knowledge in this respect.

Now, this might seem to be the merest quibble. Even if the word 'certainty' is not actually used, one might reasonably think it clear from the overall tenor of Hilbert's work that certainty was what he was in fact after. However - and this is the more important point - I think it is possible to show quite decisively that this is not so. This insistence on seeing Hilbert as in search of certainty, in my opinion, causes serious distortions in an account of Hilbert's thought. In particular, it is the root of much misunderstanding of the nature and importance of Hilbert's foundational programme.

A very striking example of such a misunderstanding occurs late in Kitcher's paper, when he considers a response Hilbert might make to an objection he has raised to Hilbert's account of our knowledge of certain *general* mathematical facts. (Briefly, the objection concerns the point that intuition, qua form of perception, seems suited only to cognition of *particular* facts - I say a little more about this in **Chapter Two**.) The proffered response would fend off Kitcher's objection by integrating mathematics closely into a successful physical theory. Kitcher thinks any such option is *closed* to Hilbert because

> . . . it would concede to skepticism the crucial point that there is no sharp difference between our knowledge of mathematics and our knowledge of physical reality.[208]

The suggestion, then, is that this would be anathema to Hilbert.

---

[207] He also speaks about sceptical attacks on classical mathematics. But of course, the skepticism he has in mind is in fact the very specific and mathematically motivated objection to classical analysis (and set theory) associated with Brouwer. Hilbert did indeed think Brouwer a skeptic, and he did indeed think of his foundational programme as a defense against Brouwerian skepticism. But it is just a mistake to infer from this any interest in much of what a philosopher might think of as skepticism about classical mathematics.
[208] Kitcher [op cit] p113.

I find this really very surprising. One of the single most striking features of Hilbert's cast of mind, evidenced throughout his writings and indeed throughout his career as a professional mathematician and administrator of a major research center in mathematical physics, was his lifelong insistence on the importance, for both disciplines, of the very closest integration of pure mathematics and theoretical physics.[209] It is well known that Hilbert, throughout his whole life, insisted repeatedly that geometry was at once part of pure mathematics and the 'most perfect' part of theoretical physics. He frequently insists that there is no clear distinction to be drawn between pure and applied mathematics, a perspective that came to be very closely associated with Göttingen mathematics. Indeed, Hilbert's efforts to reform the German mathematics curriculum in such a way as to undermine any suggestion of a principled distinction between pure and applied mathematics brought him into an acrimonious and much publicized dispute with the applied mathematician von Mises. If it is skepticism to see no sharp distinction between mathematics and physical science, then there is no doubt whatsoever that Hilbert was a sceptic. So far from combatting this kind of 'skepticism', Hilbert's Programme is in part a defense of it.

It seems to me, then, that it really is an implausibly heavy-handed interpretation which inflates a mathematician's philosophically innocent talk of the clarity and reliability of his discipline into an endorsement of some kind of controversial mathematical epistemology. Even a professional epistemologist might very well accord mathematics this status and more, and still shrug her shoulders at the question, asked in that familiar and ponderously philosophical tone, whether any mathematical truths really are *certain.*

A further consequence of this Kantian, certainty-seeking interpretation of Hilbert is to make the incompleteness results seem particularly damning. Ever so often, the import of the Second Incompleteness Theorem is said to be that a consistency proof for any mathematical theory must make use of techniques which are 'more dubitable' than those to be found within the theory with respect to which consistency is proved, and this, it is suggested, is a victory for the sceptic Hilbert was attempting to overcome.[210] Hilbert, it is said, cannot defeat the sceptic, since Gödel has shown that the success of his programme requires the

---

[209] See the last two sections of **Chapter One**. At a more anecdotal level, witness for example his delight at the literal integration of the physics and mathematics departments around what is now known as the Hilbert Space in the research buildings designed to his specifications at Göttingen.

[210] This is a familiar theme in popular writing on the incompleteness theorems, of course, but it there are at least echoes of it to be heard in serious philosophical discussions of mathematics as well.

acceptance of principles stronger, more dubitable, than those the sceptic has called into question.

But this perspective simply will not lie down with what Hilbert actually says. At best, it is a one-sided and incomplete account of what Hilbert's Programme was intended to achieve, and what it leaves out is precisely the aspect of Hilbert's thought which is of the greatest philosophical interest. Correcting this false perspective is one of the principal tasks I have undertaken in this essay.

APPENDIX TWO:
Notation, Systems of Arithmetic, and some Standard Facts

In this **Appendix**, I explain the notational conventions and document the standard facts - largely facts about formal theories of arithmetic and recursive functions - which I have taken for granted in the rest of this essay. My usage is entirely conventional, and the standard facts are restricted to those proved in every text that deals with these matters. Readers who are familiar with the relevant literature will find my usage unidiosyncratic.

(A) Notation and Syntax. For the most part, the logical and mathematical symbols employed are those of Schoenfield [1967], used in the manner explained therein. Sometimes I use '$x^{\rightarrow}$' as an abbreviation for '$x_1, \ldots, x_n$'.

Apart from this appendix, the logical symbols that appear in this text are used, not mentioned. In discussing the language of some formulation of the first order predicate calculus with identity, for example, I use the symbol '$\forall$' to denote the universal quantifier of that language, whatever it might be. In discussing the language of arithmetic, I use the symbol '+' to denote the addition function symbol of that language, whatever it might be. Similarly, I use the expression '$(\exists x)(\varphi \rightarrow \psi)$' (for instance) to refer to the object language formula that begins with the left bracket, followed by the existential quantifier, followed by . . . , followed by the right bracket. Since I shall have no occasion to discuss the visual properties the symbols in the object language, I rarely need to mention them.

My usage of raised corners is a (standard) extension of the usage explained in Quine [1951a]. According to Quine, the expression '$\ulcorner \varphi = \psi \urcorner$' abbreviates 'the result of putting φ and ψ respectively in the blanks of '... = ...'. Equivalently, the same expression abbreviates 'the result of writing φ followed by '=' followed by ψ', where the Greek letters are syntactic variables. My extension of this usage is motivated by my extensive use of code numbering. I use '$\ulcorner \varphi \urcorner$' to refer to the *code number* of the expression φ. That this is indeed an extension of the Quinean usage follows from two facts, established in Gödel [1931] and Quine [1946] respectively. Firstly (Gödel), the theory of syntax can be reduced to elementary number theory: secondly (Quine) elementary number theory can be reduced to the theory of syntax. Elementary number theory and the theory of syntax are therefore interreducible. They are mathematically equivalent.

A further piece of notation (this time attributable to Feferman - see Feferman [1960]) will help to make this clearer. I shall underline the symbol for a logical operation (on formulas) in order to denote the corresponding arithmetical operation (on code numbers): then '$\underline{\rightarrow}$' denotes the arithmetical operation that corresponds to the logical operation of (material) implication. Thus '$\ulcorner\varphi\urcorner \underline{\rightarrow} \ulcorner\psi\urcorner$' denotes the number that is the value of the arithmetical operation $\underline{\rightarrow}$ for the arguments $\ulcorner\varphi\urcorner$ and $\ulcorner\psi\urcorner$. Amongst many other things, Gödel [1931] proves that, in any theory in which the recursive functions are representable (see below), the encoding of syntax can be carried out in such a way as to ensure that the identities

$(\ulcorner\varphi\urcorner \underline{\wedge} \ulcorner\psi\urcorner) = \ulcorner(\varphi \wedge \psi)\urcorner$
$(\ulcorner\varphi\urcorner \underline{\vee} \ulcorner\psi\urcorner) = \ulcorner(\varphi \vee \psi)\urcorner$
$(\ulcorner\varphi\urcorner \underline{\rightarrow} \ulcorner\psi\urcorner) = \ulcorner(\varphi \rightarrow \psi)\urcorner$
$(\underline{\neg}\ulcorner\varphi\urcorner) = (\ulcorner\neg\varphi\urcorner)$
$(\underline{\exists}x)\ulcorner\varphi\urcorner = \ulcorner(\exists x)\varphi\urcorner$

along with such operations as concatenation, and

$\ulcorner\varphi\urcorner[v/t]$ = the result of substituting the term t for free occurrences of the variable v in φ

for example, are provable in **PRA**.

Thus, given a reasonable coding, the expression I display as

(a) $\mathbf{PRA} \vdash (\ulcorner\varphi\urcorner \underline{\wedge} \ulcorner\psi\urcorner) = \ulcorner(\varphi \wedge \psi)\urcorner$

asserts truly: It is provable in **PRA** that the number that is the value of the function $\underline{\wedge}$ for the arguments $\ulcorner\varphi\urcorner$ and $\ulcorner\psi\urcorner$ is the same as the number $\ulcorner(\varphi \wedge \psi)\urcorner$.

Given these properties of a suitable encoding, there is nothing that needs to be said about the expressions of the various languages we shall be discussing that cannot be said just as well in terms of their codes. We can, in a sense, 'identify' expressions with their code numbers. Then '$\ulcorner(\forall x)(F(x) \rightarrow G(x))\urcorner$', for example, is a singular term denoting a number, but we may equally well take it - the expression displayed between the single

quotation marks, that is - to denote the expression that is coded by that number. Sometimes I shall say things like

(b) $\ulcorner(\forall x)(x + 0 = x)\urcorner$ is derivable in **PA**

and this can seem strange, since $\ulcorner(\forall x)(x + 0 = x)\urcorner$ is a number, rather than a formula. But the impression of strangeness will pass if we remember that formulas are being 'identified' with their code numbers. If it helps, one may think of (b) as akin to (c)

(c) Harry is derivable in **PA**

where 'Harry' is a name that has been allocated to a formula in the language of **PA** - the formula that begins with the left bracket, followed by the universal quantifier . . . etc.

This talk of 'identifying' formulas with their code numbers should not be taken too seriously, however. In particular, the suggestion that code numbers are being used as names of formulas should be treated with caution. We arithmetize syntax in order to bring to bear upon the study of syntax a rich body of results from elementary number theory. When this was first done, in Gödel [1931], little was known about syntax, but a great deal was known about elementary number theory. By defining a *purely syntactic operation on expressions*, carrying expressions from the language of arithmetic to expressions in a syntax language, Gödel was able to bring this rich body of knowledge to bear upon the study of syntax. Syntactic operations, such as that of substituting a constant for free occurrences of a variable, are made to correspond to arithmetical operations on their codes. One can, if one wishes, put this in a semantic mode, by saying (for example) that truths of elementary number theory are turned into truths of formal syntax. But this should not be taken to betoken a richer semantic relation that this: the theory of syntax and elementary number theory are interreducible.

I use boldface type in the text in three ways. Firstly, I use '**PRA**' (for example) to denote some standard formalization of primitive recursive arithmetic, and '**ZF**' to denote some standard formalization of Zermelo-Frankel set theory. '**PRA**', then, refers to some axiom system and its associated language. 'PRA', on the other hand, refers to primitive recursive arithmetic.

Secondly, I use boldface type to indicate the numeral for some particular number, or in general, the syntactic object corresponding to some mathematical constant, predicate, operation, or whatever. Thus '**0**' denotes the standard numeral for zero. Similarly, I use " ' " to denote the symbol for the successor function, and ' + ' to denote the symbol for the addition function. Boldfacing therefore distinguishes 0 from **0** (for example).

Notice that '$\ulcorner \mathbf{n} \urcorner$', therefore, denotes the code number of the numeral for the number n.

Thirdly, I use '**1**' as an *abbreviation* of " **0**' ", which is the standard numeral for the number one. Similarly, I use '**2**' as an abbreviation of " **0**'' " , . . . , '**n**' as an abbreviation of " **0**'···' ", with n iterations of " ' ".

(B) Languages. A *language* in this essay is a set of constants. The language of set theory, for example, is $\{\in\}$. Particularly important for our purposes is the language L(A) = $\{\mathbf{0}, ', +, \times\}$, the *language of arithmetic*. Under the intended interpretation of L(A), these constants are interpreted by the natural number zero, the successor function, the addition function, and the multiplication function respectively.

Languages are interpreted by specifying an interpretation function. In the case of a mathematical language, therefore, the interpretation of a language is itself a mathematical object, which we may take to be a function from natural numbers (identifying expressions with their codes) to, say, sets. Typically, there will be an *intended* interpretation of a mathematical language. In the case of L(A), the intended interpretation is the one described above. In the case of $\{\in\}$, the intended interpretation assigns to $\in$ the relation of set-membership. The intended interpretation is given in English, or a mathematical-looking extension thereof, with all the attendant foibles of communication in natural languages. For all that, it is still a mathematical object that is being given.

(C) Systems of Arithmetic and Some Standard Facts. I use the words 'system' and 'theory' interchangeably. Normally, a *theory* is a set of sentences closed under logical consequence. Sometimes, however, the logical facts demand a more fine-grained notion of theory, on which different theories pick out the same set of sentences - the facts associated with the Second Incompleteness Theorem in particular make this demand. I discuss this further in **Chapter Three**.

A sentence S of L(A) is *true* if it is true in the intended model of L(A), as described above.

A (number-theoretic) function is a *recursive function* if it is included in the smallest class containing the initial functions (the zero function, the projection functions, and the successor function) and closed under the operations of composition, primitive recursion, and minimization of regular functions. A (number theoretic) function is called *primitive recursive* if it is included in the smallest class containing the initial functions, and closed under the operations of composition and primitive recursion. A relation is *recursive* if it has a characteristic function which is recursive, and a set of numbers is *recursively enumerable* if it is either empty or the range of some (total) recursive function of one variable (roughly, if it is the output of some computing machine).

It will be useful to have a measure of the quantificational complexity of formulas. An *atomic formula* (of L(A)) is an equation t = t', where t, t' are terms which need not be closed. A formula (of L(A)) is called a *bounded formula* if it belongs to the smallest class containing all equations of L(A), and containing $\neg$F, (F $\wedge$ G), (F $\vee$ G), ($\forall$x<y F), and ($\exists$x<y F) whenever it contains F and G. The bounded formulas are also known as Σ_0 *formulas* or Π_0 *formulas* indifferently. From an Σ_n formula , one obtains a Π_{n+1} *formula* by prefixing 0 or more universal quantifiers. From a Π_n formula, one obtains a Σ_{n+1} *formula* by prefixing 0 or more existential quantifiers. A relation (on ω) is called a Σ_n *relation* (Π_n relation) if it is the extension of a Σ_n formula (Π_n formula). A relation (on ω) is called a Δ_n *relation* if it is both Σ_n and Π_n.

The Σ_n relations are closed under unions, finite intersections, bounded quantifications, and unbounded existential quantifications. The Π_n relations are closed under unions, finite intersections, bounded quantifications, and unbounded universal quantifications. The Π_n relations are the complements of the Σ_n relations. It can be shown that *the Σ_n relations are the recursively enumerable relations*, whilst the Δ_n relations are the recursive relations.[211]

---

[211] See Tarski, Mostowski, and Robinson [1953] p56 ff.

An n-place function $f(x_1, \ldots, x_n)$ is *representable* in a theory **T** if there is a formula $F(x_1, \ldots, x_n, x_{n+1})$ of L(**T**) such that for any n+1 natural numbers $n_1, \ldots, n_n, k$: if $f(x_1, \ldots, x_n) = k$, then $\mathbf{T} \vdash (\forall x_{n+1})(F(n_1, \ldots, n_n, x_{n+1}) \leftrightarrow x_{n+1} = k)$. An n-place relation S is *weakly representable* in **T** by the formula φ if, for any n natural numbers $k_1, \ldots, k_n$, $\langle k_1, \ldots, k_n \rangle \in S$ iff $\mathbf{T} \vdash \varphi(\mathbf{k_1}, \ldots, \mathbf{k_n})$. An n-place relation S is *strongly representable* in **T** (or *numeralwise expressible* in **T**) by the formula φ if S is weakly representable by φ in **T** and, for any n natural numbers $k_1, \ldots, k_n$, $\langle k_1, \ldots, k_n \rangle \notin S$ iff $\mathbf{T} \vdash \neg\varphi(\mathbf{k_1}, \ldots, \mathbf{k_n})$.

Except for the system $\mathbf{Z_2}$, all of the following systems presented as theories with standard formalization (in the sense of Tarski, Mostowski and Robinson).[212] The first two, **Q** and **R**, along with **PA** itself, are theories in the language of arithmetic L(A).

Our first theory, the system **R** of *Robinson's Arithmetic*, has the following seven proper axioms:

(R1) $\mathbf{n} + \mathbf{p} = \mathbf{n + p}$.
(R2) $\mathbf{n} \times \mathbf{p} = \mathbf{n \times p}$
(R3) $\mathbf{n} \neq \mathbf{p}$ if $n \neq p$
(R4) $(\forall x)(x < \mathbf{n} \vee x = \mathbf{n} \vee \mathbf{n} < x)$
(R5) $(\forall x)(\forall y)(x < y \leftrightarrow (x \neq y \wedge (\exists z)(z + x = y)))$
(R6) $(\forall x) \neg x < 0$.
(R7) $(\forall x)(x < \mathbf{n + 1} \rightarrow (x = \mathbf{0} \vee x = \mathbf{1} \vee \ldots \vee x = \mathbf{n}))$

The system **Q** - which is often referred to as Robinson's arithmetic also - has the following seven proper axioms:

(Q1) $(\forall x)(\forall y)(x' = y' \rightarrow x = y)$
(Q2) $(\forall x)(0 \neq x')$
(Q3) $(\forall x)(x \neq 0 \rightarrow (\exists y)(x = y'))$
(Q4) $(\forall x)(x + 0 = x)$
(Q5) $(\forall x)(\forall y)(x + y' = (x + y)')$
(Q6) $(\forall x)(x \cdot 0 = 0)$

---

[212] See Tarski, Mostowski, and Robinson [1953] pp52-53. In Schoenfield [1967], $\mathbf{Z_2}$ is presented in standard formalization.

(Q7) $(\forall x)(\forall y)(x \cdot y' = (x \cdot y) + x)$

R is important because of the following facts:

> If S is a true Σ_1 sentence, then S is *provable* in **R**. If F is a recursively enumerable relation, then F is *weakly representable* in **R**. If F is a recursive relation, then F is *strongly representable* (or numeralwise expressible) in **R**. If f is a recursive function, then f is *representable* in **R**.[213]

These facts hold in extensions of **R**, such as **Q**.[214]

The system **PRA** can be obtained from Q by deleting axioms **(Q4)** - **(Q7)**, and substituting axioms corresponding to the definitions of all the primitive recursive functions. Equivalently, we can follow Robbins [1969] and substitute for **(Q4)** - **(Q7)** the following axiom schemas:

(Q4') $I_i^n(t_1, \ldots, t_n) = t_i$
(Q5') $Rgh(t_1, \ldots, t_n, 0) = g(t_1, \ldots, t_n)$
(Q6') $Rgh(t_1, \ldots, t_n, t_0') = h(t_1, \ldots, t_n, t_0, Rgh(t_1, \ldots, t_n, t_0))$
(Q7') $Chg_1, \ldots, g_m(t_1, \ldots, t_n) = h(g_1(t_1, \ldots, t_n), \ldots, g_m(t_1, \ldots, t_n))$

(where g, h are already given (and therefore *primitive recursive*) functions, I_i^n is the n place projection function, and R and C are the primitive recursion and composition operations respectively), along with the induction schema

(Ind) $[A(0) \wedge (\forall x)(A(x) \rightarrow A(x'))] \rightarrow (\forall x)A(x)$

Alternatively, we can give up standard formalization, and follow Goodstein [1971] by setting out PRA in the form of a 'logic free' equation calculus. Here, we let A, B be any recursive terms (recursive functions or numerals), and take as axioms all primitive recursive definitions. With F, G any recursive functions, ' the successor function, + the addition function, P the predecessor function, - the monus (cut-off subtraction) function, and x any variable, the inference rules are the following:

---

213 See Tarski, Mostowski, and Robinson [1953] pp56 ff.
214 See Tarski, Mostowski, and Robinson [loc cit].

(Equality) $A = B$
$\underline{A = C}$
$B = C$

(Subst 1) $\underline{F(x) = G(x)}$
$F(A) = G(A)$

(Subst 2) $\underline{A = B}$
$F(A) = B(A)$

(Uniqueness 1) $\underline{F(x) = F(x')}$
$F(x) = F(0)$

(Uniqueness 2) $\underline{F(x) = F(x)'}$
$F(x) = F(0) + x$

(Uniqueness 3) $\underline{F(x) = PF(x)}$
$F(x) = F(0) - x$

(Uniqueness 4) $F(0) = G(0)$
$\underline{F(x') = G(x')}$
$F(x) = G(x)$

Goodstein then shows how to define the sentential operators from these equations.[215] He also shows that anything derivable in the standard formalization of PRA given above by use of the induction schema can be derived in this equation calculus, by associating with the premises P(0) and $P(x) \rightarrow P(x')$ the equations $p(0) = 0$ and $(1 - p(x))p(x') = 0$ respectively (where p is the equation associated with the predicate P), and showing that the equation $p(x) = 0$ associated with the conclusion P(x) is then derivable.[216]

Goodstein's 'logic free' formalization helps make it clear that PRA permits induction *only with respect to predicates definable by composition and primitive recursion.* We owe to Tait [1981] a compelling defense of the view that PRA is the most plausible candidate interpretation of Hilbert's finitary mathematics. By finitary arguments, then, I shall mean arguments formalizable in **PRA**. By finitary proofs, I shall mean proofs in **PRA**. If we then add to **(Q1)** - **(Q7')** plus **(IND)** an axiom schema corresponding to the *minimization* operation, the result is be a system equivalent to full Peano Arithmetic. However, the standard formalization **PA** of Peano Arithmetic is obtained by adding to **(Q1)** - **(Q7)** all instances of the schema **(Ind)**, *without* restriction on A(x).[217] **PA** is not plausibly

---

[215] Goodstein [1971] pp120-121.
[216] Goodstein [op cit] pp121-122.
[217] You will notice that the system Q adds to three axioms giving the uniqueness of zero, and the existence and uniqueness of successors, the recursion equations for plus and times. **PRA**, in effect, extends this process by adding equations defining all the primitive recursive equations. However, it is possible to regard these equations as fixing interpretations for denumerably many of the function signs available in the

regarded as a finitary system in Hilbert's sense, primarily in virtue of the availability of induction for non-primitive recursive predicates.

The final system we shall need to mention is the system $\mathbf{Z_2}$ of *second-order arithmetic*. As formalized in second-order logic, the background logic makes available all instances of the comprehension schema

(Comp) $(\exists X^n)(\forall x_1) \ldots (\forall x_n)(X^n(x_1, \ldots, x_n) \leftrightarrow \varphi(x_1, \ldots, x_n))$

(where X^n is an n-place relation symbol, $x_1, \ldots, x_n$ are individual variables, and $X^n(x_1, \ldots, x_n)$ does not occur free in φ). The proper axioms of $\mathbf{Z_2}$ are those of **PA**, except that the axiom schema of induction (**A8**) is replaced by the second order sentence

(A8') $(\forall A)[(A(0) \wedge (\forall x)(A(x) \rightarrow A(x'))) \rightarrow (\forall x)A(x)]$.

As is pointed out in Hilbert and Bernays [1939], Supplement Four, $\mathbf{Z_2}$ is a system in which all of classical analysis (and much more) can be adequately formalized.[218]

In the sequence of theories **R**, **Q**, **PRA**, **PA**, and $\mathbf{Z_2}$, each theory is a subtheory of the following theory. Since the recursive functions are all representable in **R**, they are all representable in the theories of which **R** is a subtheory.

---

background logic, and therefore as belonging in the background language, rather than part of the axiomatic content of PRA. I choose this description of PRA primarily to facilitate comparisons between **PRA**, **Q**, and **PA**.

[218] See also Simpson [1988] pp350-351. In Simpson's specification of the Conservation Programme, its objective becomes a proof in **PRA** that $\mathbf{Z_2}$ is conservative over **PRA** with respect to Π_1 sentences.

References

Ackermann, W: - [1940] 'Zur Widerspruchsfreiheit der Zahlentheorie', *Mathematische Annalen* **117** (1940) pp162-194.

Andrews, P.B: - [1986] *An Introduction to Mathematical Logic and Type Theory: to Truth through Proof* (London 1986).

Appel, K.I., and Haken, W: - [1976] 'Every Planar Map is Four Colorable', in *Bulletin of the American Mathematical Society*, **82** (1976) pp711-712.

Asprey, W., and Kitcher, P. (eds): - [1988] *Minnesota Studies in the Philosophy of Science vol XI: History and Philosophy of Modern Mathematics* (Minneapolis, 1988).

Auerbach, D: - [1985] 'Intensionality and the Gödel Theorems', *Philosophical Studies* **48** (1985) pp337-351.

Barwise, J. (ed): - [1977] *Handbook of Mathematical Logic* (Amsterdam 1977).

Benacerraf, P., and Putnam, H: - [1964] *Philosophy of Mathematics: Selected Readings* (New York 1964].

[1983] *Philosophy of Mathematics: Selected Readings*, second edition (New York 1983).

Bernays, P: - [1934] 'On Platonism in Mathematics', English translation in Benacerraf and Putnam (eds) [1983] pp258-271.

[1942] Review of M. Steck, *Ein unbekannter Brief von Gottlob Frege Uber Hilberts ersteVorlesung über die Grundlagen der Geometrie*, in *JSL* 7 (1942)pp92-93.

[1967] 'David Hilbert', in P. Edwards (ed) [1967].

Blumenthal, O: - [1035] 'Lebensgeschichte', in Hilbert [1935], pp388-429.

Boolos, G: - [1971] 'The Iterative Concept of Set', *JPhil* **68**, reprinted in Benacerraf and Putnam, eds. [1964], pp486-582.

[1979] *The Unprovability of Consistency: An Essay in Modal Logic* (Cambridge 1979).

[1987] 'A Curious Inference', *Journal of Philosophical Logic* **16** (1987) pp1-12.

Boolos, G., and Jeffrey, R: - [1989] *Computability and Logic* (Cambridge 1989).

Cantor, G: - [1955] *Contributions to the Founding of the Theory of Transfinite Numbers* (New York 1955).

Chihara, C: - [1982] 'A Gödelian Thesis Regarding Mathematical Objects', *Phil Review* **91** (1982) pp211-227.

[1990] *Constructibility and Mathematical Existence* (Oxford 1990).

Church, A: - *Introduction to Mathematical Logic, vol 1* (Princeton 1956).

Churchland, P., and Hooker, C.A. (eds): - [1985] *Images of Science* (Chicago 1985).

Davis, P.J., and Hersh, R: - [1980] *The Mathematical Experience* (Harmonsworth 1980).

Dedekind, R: - [1887] 'Was sind und was sollen die Zahlen?', trans. as 'The Nature and Meaning of Number' in Dedekind [1963].

[1872] 'Stetigkeit und irrationale Zahlen', trans. as 'Continuity and Irrational Numbers' in Dedekind [1963].

[1963] *Essays on the Theory of Numbers* , trans. Beman, W.W. (New York 1963). All references to Dedekind are to this edition.

Detlefsen, M: - [1986] *Hilbert's Program* (Dordrecht 1986).

Dreben, B., Andrews, P. B., and Anderaa, S: - [1963] 'False Lemmas in Herbrand', *Bulletin of the American Mathematical Society* **69** (1963) pp699-706.

Edwards, H: - [1988] 'Kronecker's Place in History', in Asprey, W., and Kitcher, P. (eds) [1988] pp139-145.

Edwards, P: - [1967] *The Encyclopedia of Philosophy* (New York 1967).

Epstein, R.L., and Carnielli, W.A: - [1989] *Computability: Computable Functions, Logic, and the Foundations of Mathematics* ((Belmont CA 1989).

Fang, J: - [1970] *Hilbert: Towards a Philosophy of Modern Mathematics II* (New York 1970).

Feferman, S: - [1960] 'Arithmetization of Syntax in a General Setting', *Fundamentae Mathematicae* **49** (1960) pp35-92.

[1964] 'Systems of Predicative Analysis I, II', *JSL* **29** (1964) pp1-30.

Field, H: - [1980] *Science Without Numbers* (London 1980).

Fine, A: - [1986] *The Shaky Game* (Chicago, 1986).

Fodor, G: - [1981] *Representations* (Brighton 1981)

Frege, G: - [1979] *Posthumous Writings* ed. Hermes, H., Kambartel, F., and Kaulbach, F., trans. by Long, P., and White, R. (Oxford 1979)

[1980] *Philosophical and Mathematical Correspondence*, ed. McGuinness, B., trans Kaal, H. (Oxford 1980).

Friedman, H: - [1976] 'Systems of second order arithmetic with restricted induction I, II' (abstracts), *JSL* **41** (1976) pp557-559.

Gauss, C.F: - [1880] *Carl Friedrich Gauss - Friedrich Wilhelm Bessel: Briefwechsel* (Hildesheim 1975)

Geach, P., and Black, M: - [1952] *Translations from the Philosophical Writings of Gottlob Frege* (Oxford 1952).

Genzen, G: - [1936] 'Die Widerspruchsfeiheit der reinen Zahlentheorie', *Mathematische Annalen* **112** (1936) pp493-565, English translation in Genzen [1969] pp132-213.

[1938] 'Die gegenwärtige Lage in der mathematischen Grundlagenforschung', *Forschunger zur Logik und zur Grundlegung der exakten Wissenschaften*, New Series **4** (Leipzig 1936), trans. in Genzen [1969] pp234-251.

[1969] *The Collected Papers of Gerhard Genzen*, ed. C. Szabo (Amsterdam 1969). All references are to this edition.

Gödel, K: - [1931] 'Über formal unentscheidbare Sätze der *Principia mathematica* und verwandter Systeme 1', *Monatshefte für Mathematik und Physik* 38 (1931) pp173-198, reprinted with facing English translation in Gödel [1986] pp144-195. All references are to this edition.

[1933] 'On intuitionistic arithmetic and number theory', in Gödel [1986] pp287-295.

[1964] 'What is Cantor's Continuum Problem', revised version. All references are to the translation in Gödel [1990].

[1986] *Collected Works, vol I* (Oxford 1990).

[1990] *Collected Works, vol II* (Oxford 1990).

Goldfarb, W: - [1979] 'Logic in the Twenties: the Nature of the Quantifier', *JSL* **44** (1979) pp351-368.

[1990] 'Herbrand's Theorem and the Incompleteness of Arithmetic', *Iyyun* **39** (1990) pp45-64.

Goodman, N., and Quine, W.V.O: - [1947] 'Steps Towards a Constructive Nominalism', *JSL* **12** (1947) pp97-122.

Goodstein, R.L: - [1971] *Development of Mathematical Logic* (New York 1971).

Gray, J: - [1989] *Ideas of Space*, second edition (Oxford, 1989).

Hallett, M: - [1990] 'Physicalism, Reductionism, and Hilbert', unpublished draft.

[1989] *Mathematics and Mind: an Essay on Kant and Hilbert* (unpublished manuscript).

Herbrand, J: - [1930] *Recherches sur la théorie de la démonstration* (PhD Thesis, University of Paris 1930, English translation in Herbrand [1971]).
[1971] *Logical Writings* , ed. W. Goldfarb (Cambridge, Mass. 1971).

Hilbert, D: - [1900] 'Mathematische Probleme', *Göttinger Nachrichten* (1900) pp253-297. Refs. are to the English translation in *Proceedings of Symposia in Pure Mathematics* **28** (1976), pp34.
[1904] 'Über die Grundlagen der Logik und der Arithmetik', *Verhandlungen des Dritten Internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904* (Leipzig 1905). English translation in van Heijenoort (ed) [1967] pp128-138.
[1917] 'Axiomatisches Denken', *Mathematische Annalen* **78** (1918). Refs. are to the English translation in Fang [1970].
[1925] 'Über das Unendliche', *Mathematische Annalen* **95** (1926) pp161-190. Refs. are to the English translation in van Heijenoort [1967].
[1927] 'Die Grundlagen der Mathematik', *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Univerität 6* (Leipzig, 1927). Refs. are to the English translation in van Heijenoort (ed) [1967].
[1935] *Gesammelte Abhandlungen, dritte Band* (Berlin 1935).
[1971] *The Foundations of Geometry* (LaSalle 1971). Second English edition of *Die Grundlagen der Geometrie*, trans by Unger, L., from the Tenth German Edition.

Hilbert, D., and Bernays, P: - [1934] *Die Grundlagen der Mathematik , vol I* (Berlin 1934).
[1939] *Die Grundlagen der Mathematik, vol II* (Berlin 1939).

Hilbert, D., and Cohn-Vossen, S: - [1952] *Geometry and the Imagination* (New York 1952). English translation by P. Nemenyi of *Anschauliche Geometrie.*

Hunter, G: - [1980] 'What do the Consistency Proofs for non-Euclidean Geometries Prove',*Analysis* **40** (1980) pp79-83.

Jech, T: - [1978] *Set Theory* (London 1978).

Jeroslow, R: - [1973] 'Reduncancies in the Hilbert-Bernays derivability conditions for Gödel's Second Incompleteness Theorem', *JSL* **38** (1973) pp359-367.

Kessler, G: - [1978] 'Mathematics and Modality', *Noûs* **12** (1978) pp421-441.

Kitcher, P: - [1975] 'Kant and the Foundations of Mathematics', *Phil Review* **84** (1975) pp23-50.

[1976] 'Hilbert's Epistemology', *Philosophy of Science* **43** (1976), pp99-115.

Kleene, S.C: - [1952] *An Introduction to Metamathematics* (Princeton, 1952).

Kochen, S., and Kripke, S: - [1981] 'Non-standard models of Peano arithmetic', *L'enseignement mathematique*, Second Series **28** (1981) pp211-231.

Kreisel, G: - [1964] 'Hilbert's Programme', in Benacerraf, P., and Putnam, H. (eds) [1964].

[1965] 'Mathematical Logic', in T.L. Saaty (ed) [1965].

Kreisel, G., and Takeuti, G: - [1974] 'Formally Self-Referential Propositions for Cut-Free Classical Analysis and Related Systems', *Dissertationes Mathematicae* **118** (1974) pp4-50.

Kronecker, L: - [1887] *Werke: Band III* (Leipzig 1887).

Leisenring, A.C: - [1969] *Mathematical Logic and Hilbert's ε-Symbol* (New York 1969).

Löb, M: - [1955] 'Solution to a problem of Leon Henkin', *JSL* **20** (1955) pp115-118.

Luce, L: - [1989] 'Platonism from an Empirical Point of View', *Philosophical Topics* **17** no 2 (1989), pp109-128.

Mach, E: - [1960] *The Development of Mechanics* , trans. McCormack, T.J., (LaSalle 1960), from *Die Mechanik in ihrer Entwicklung historisch-kritisch dargestellt* (Leipzig 1912).

Maddy, P: - [1988a] 'Believing the Axioms 1', *JSL* **53** no 2 (1988).

[1988b] 'Believing the Axioms 2', *JSL* **53** no 3 (1988).

[1990] *Realism in Mathematics* (Oxford, 1990).

McGee, V: - [1991] *Truth, Vagueness, and Paradox* (Indianapolis 1991).

Monk, D: - [1976] *Mathematical Logic* (New York 1976).

Montague, R: - [1962] 'Theories incomparable with respect to relative interpretability', **JSL 27** (1962) pp195-211.

Moore, A.W: - [1990] *The Infinite* (London 1990).

Moore, G: - [1982] *Zermelo's Axiom of Choice* (Berlin 1982).

Mostowski, A: - [1952] 'On models of axiomatic systems', *Fundamentae Mathematicae* **39** (1952) pp133-158.

[1966] *Thirty Years of Foundational Studies* (London 1966).

Parsons, C: - [1971] 'Ontology and Mathematics', *Phil Review* **80** (1971) pp151-176, reprinted in Parsons [1983].

[1980] 'Mathematical Intuition', *PAS* (1980) pp145-168.

[1983] *Mathematics in Philosophy* (Ithaca, NY 1983).

Plantinga, A: - [1974] *The Nature of Necessity* (Oxford 1974).

Pohlers, W: - [1989] *Proof Theory: an Introduction* (Berlin 1989).

Prawitz, D: - [1981] 'Philosophical Aspects of Proof Theory', in *Contemporary Philosophy: a New Survey vol I* pp235-277 (London 1981).

Putnam, H: - [1979] *Mathematics, Matter, and Method - Philosophical Papers vol I* (second edition, Cambridge 1979).

Quine, W.V.O: - [1946] 'Concatenation as a basis for arithmetic', *JSL* **11** (1946) pp105-114.

[1951] *From a Logical Point of View* [Cambridge, Mass. 1951]

[1951a] *Mathematical Logic* , revised edition (Cambridge, Mass. 1951).

[1960] *Word and Object* (Cambridge, Mass. 1960).

[1969] *Ontological Relativity and Other Essays* (New York 1969).

[1970] *Philosophy of Logic* (Englewood 1970).

[1974] *The Roots of Reference* (Cambridge, Mass. 1974).

[1984] Review of Parsons [1983], *JPhil* **81** (1984) pp783-794.

[1990] *Truth* (Cambridge, Mass. 1990)

Reid, C: - [1986] *Hilbert/Courant* (New York 1986).

Resnik, M: - [1974] 'The Frege-Hilbert Controversy', *Philosophy and Phenomenological Research* **34** (1974) pp386-403.

[1974a] 'On the Philosophical Significance of Consistency Proofs', *Journal of Philosophical Logic* **3** (1974) pp133-147. References are to the reprint in S. Shanker (ed) [1988].

Robbin, J: - [1969] *Mathematical Logic* (New York (1969).

Rosser, J.B: - [1939] 'An informal exposition of proofs of Gödel's theorem and Church's theorem', *JSL* **4** (1939) pp53-60.

Saaty, T.L. (ed): - [1965] *Lectures on Modern Mathematics*, vol III (New York 1965).

Saaty, T.L., and Kainen, P.C: - [1977]*The Four-Color Problem: Assaults and Conquest* (New York 1977).

Scanlon, T: - [1973] 'The consistency of number theory via Herbrand's theorem', *JSL* **38** (1973) pp29-58.

Schoenfield, J: - [1967] *Mathematical Logic* (Reading, Mass. 1967).

Scott, D: - [1974] 'Axiomatizing Set Theory', *Proceedings of Symposia in Pure Mathematics* **13**, Part II pp207-214.

Shanker, S. (ed): - [1988] *Gödel's Theorem in Focus* (New York 1988).

Sieg, W: - [1988] 'Hilbert's Programme Sixty Years Later', *JSL* **53** no 2 (1988), pp338-348.

Simpson, S: - [1988] 'Partial Realizations of Hilbert's Programme', *JSL* **53** no. 2, June 1988 pp349-363.

Skolem, T: - [1923] 'The foundations of elementary arithmetic' , English trans. in van Heijenoort (ed) [1967]

Smart, J.R: - [1973] *Modern Geometries*, third edition (Belmont CA 1973).

Smorynski, C: - [1977] 'The Incompleteness Theorems', in Barwise, J. (ed) [1977].
[1981] 'Fifty Years of Self-Reference in Arithmetic', *Notre Dame Journal of Formal Logic* **22** no 4 (1981) pp357-374.
[1985] *Self Reference and Modal Logic* (New York 1985).

Smullyan, A: - [1961] *Theory of Formal Systems* (Princeton 1961).

Steiner, M: - [1975] *Mathematical Knowledge* (Ithaca 1975).

Tait, W.W: - [1965] 'The Substitution Method', *JSL* **30** (1965) pp175-192.
[1981] 'Finitism', *JPhil* **78** (1981) pp524-546.
[1986] 'Truth and Proof: the Platonism of Mathematics', *Synthese* **69** (1986) pp341-370.

Takeuti, G: - [1987] *Proof Theory*, second edition (Amsterdam 1987).

Tarski, A., Mostowski, A., and Robinson, R: - [1953] *Undecidable Theories* (Amsterdam 1953).

van Fraassen, B: - [1980] *The Scientific Image* (Oxford 1980).

van Heijenoort, J: - [1967] *From Frege to Gödel: a Source Book in Mathematical Logic* (Cambridge Mass. 1967).

Wang, H: - [1963] *A Survey of Mathematical Logic* (Amsterdam 1963).

Wetzel, L: - [1989] 'Expressions vs. Numbers', *Philosophical Topics* **27** no 2 (1989) pp173-196.

Wright, C: - [1980] *Wittgenstein on the Foundations of Mathematics* (London 1980).

Yasuhara, A: - [1971] *Recursive Function Theory and Logic* (New York 1971).