This paper appeared in *The Journal of Philosophy* Vol. XCIII, Number 7 (July 1996) pp 317-336;
reprinted in *The Philosophers' Annual* 1998;
reprinted in *Gottlob Frege: Critical Assessments of Leading Philosophers* (Reck, E. and M. Beaney, eds), Routledge 2005.

FREGE AND HILBERT ON CONSISTENCY[1]

Patricia A. Blanchette
University of Notre Dame

Frege's work in logic and the foundations of mathematics centers on claims of logical entailment; most important amongst these is the claim that arithmetical truths are entailed by purely logical principles. Occupying a less central but nonetheless important role in Frege's work are claims about failures of entailment. Here the clearest examples are his theses that the truths of geometry are not entailed by the truths of logic or of arithmetic, and that some of them are not entailed by each other. As he, and we, would put it: the truths of Euclidean geometry are independent of the truths of logic, and some of them are independent of one another.[2]

Frege's talk of independence and related notions sounds familiar to a modern ear: a proposition is independent of a collection of propositions just in case it isn't a consequence of that collection, and a proposition or collection of propositions is consistent just in case no contradiction is a consequence of it. But some of Frege's views and procedures are decidedly un-modern. Despite developing an extremely sophisticated apparatus for demonstrating that one claim is a consequence of others, Frege offers not a single demonstration that one claim is *not* a consequence of others. Thus in particular, he gives no proofs of independence or of consistency. This is no accident. Despite his firm commitment to the independence and consistency claims just mentioned, Frege holds that independence and consistency cannot systematically be demonstrated.[3]

Frege's view here is particularly striking in light of the fact that his contemporaries had a fruitful and systematic method for proving consistency and independence, a method which was well known to him. One of the clearest applications of this method in Frege's day came in Hilbert's 1899 *Foundations of Geometry*,[4] in which Hilbert establishes via essentially our own modern method the consistency and independence of various axioms and axiom-systems for Euclidean geometry. Frege's reaction to Hilbert's work was that it was simply a failure: that its central methods were incapable of demonstrating consistency and independence, and that its usefulness in the foundations of mathematics was highly questionable.[5] Regarding the general usefulness of the method, it is clear that Frege was wrong; the last 100 years of work in logic and mathematics gives ample evidence of the fruitfulness of those techniques which grow directly from the Hilbert-style approach. The standard view today is that Frege was also wrong in his claim that Hilbert's methods fail to demonstrate consistency and independence. The view would seem to be that Frege largely missed Hilbert's point, and that a better understanding of Hilbert's techniques would have revealed to Frege their success. Despite Frege's historic role as the founder of the methods we now use to demonstrate positive consequence-results, he simply failed, on this account, to understand the ways in which Hilbert's methods could be used to demonstrate negative consequence-results.

The purpose of this paper is to question this account of the Frege-Hilbert disagreement. By 1899, Frege had a well-developed view of logical consequence, consistency, and independence, a view which was central to his foundational work in arithmetic, and to the epistemological significance of that work. Given *this* understanding of the logical relations, I shall argue, Hilbert's demonstrations do fail. Successful as they were in demonstrating significant metatheoretic results, Hilbert's proofs do not establish the consistency and independence, in Frege's sense, of geometrical axioms. This point is important, I think, both for an understanding of the basis of Frege's epistemological claims about mathematics, and for an understanding of just how different Frege's conception of logic is from the modern model-theoretic conception which has grown out of the Hilbert-style approach to consistency.

First, an account of Hilbert's method.

I. Hilbert

I.1 Syntactic Consistency

Hilbert's purpose in the *Foundations of Geometry* is to provide a complete set of axioms for geometry, to show the consistency of those axioms, and to show the independence of various axioms from others. As Hilbert conceives of them, axioms are partially-interpreted sentences; though the usual logical terms in these sentences have fixed meanings, the non-logical terms (including the geometrical 'point', 'line', 'between', etc.) are purely schematic, and have no fixed semantic value. Thus the axioms do not concern any particular subject-matter, and have no truth-values. Rather, they are open to various *interpretations*, i.e., assignments of properties and relations (or sets) to the non-logical terms. An axiom will typically be true on some interpretations, false on others.

A set of axioms is *consistent*, for Hilbert, just in case no contradiction can be deduced from it via the laws of logic. An axiom □ is *independent* of a set □ of axioms just in case □ cannot be deduced from □ via the laws of logic. Since on this account □ is independent of □ just in case the set □;~□ is consistent, Hilbert can (and does) demonstrate independence results via consistency-proofs.

A consistency-proof in the *Foundations* is always a proof of relative consistency, and takes place in two stages. Where □ is the set of axioms in question, the first stage is the construction of an interpretation for each of the non-logical terms in □. The second stage is the argument that each member of □, so interpreted, is true. This second stage makes essential use of the "background theory," i.e. the theory of those concepts, objects and relations used in the interpretation. Consider for example Hilbert's first interpretation, which assigns to 'point' the set of pairs <x, y> of numbers in a set □, to 'line' the set of ratios [u: v: w] of members of □, and to 'lies on' the set of pairs <<x, y>, [u: v: w]> such that ux+vy+w=0.[6] The demonstration that axiom

(I.2) For every two points there exists at most one line on which lie those points

is true on this interpretation is the demonstration that for each pair <a, b> of members of □, if ua+vb+w=0, and u'a+v'b+w'=0 for <u, v, w> and <u', v', w'> triples of members of □, then u:v = u':v' and v:w = v':w'. In this example, the background theory is the arithmetic of the field □.

It follows from the two steps of Hilbert's proof that no contradiction is deducible from □ unless it is also deducible from the background theory. For the background theory guarantees the truth of every sentence in □ as interpreted; hence it guarantees the truth of every deductive consequence of □ as interpreted; and contradictions, under any interpretation, are contradictions. As Hilbert puts it at the conclusion of the first consistency proof, "Every contradiction in the consequences of the [axioms] would therefore have to be detectable in the arithmetic of the field □."[7] The consistency of the set of axioms in question is reduced to the consistency of the background theory.

Assuming the consistency of the background theory, then, Hilbert's consistency-proof demonstrates that no contradiction is *deducible from* the set of axioms in question. Here there is an implicit restriction on the relation of deducibility: the principles by means of which sentences are deducible from one another must be independent of any content had by the nonlogical terms in those sentences. Otherwise, a different consistency-proof would be required for each potential interpretation, which is precisely what Hilbert wants to avoid. In line with the emerging standards of rigor of the time, Hilbert's conception of the rules of deducibility is that they are topic-neutral, sensitive only to the syntactic form of the sentences in question.

When no contradiction is deducible from a set via such syntactically-specifiable rules, we shall say that the set is *syntactically consistent.* Assuming the consistency of the background theory, then, Hilbert's consistency-proofs establish the syntactic consistency of sets of sentences.[8] The syntactic consistency of a set □;~□ implies what we shall call the *syntactic independence* of □ from □; i.e., it shows that the sentence □ is not deducible via the syntactically-specifiable laws of logic from □.

I.2 Property-Consistency and Structures

In addition to the syntactic sense of consistency just noted, in which consistency is directly tied to deducibility, there is a second sense in which a set of Hilbert-style schematic axioms can be said to be consistent. Consider for example the two sets of axioms

(□1) {(□x)(Px □ (□y)Rxy); (□x)(Px & (□y)~Ryx)}

(□2) {(□x)(Px □ (□y)Rxy); (□x)(Px & (□y)~Rxy)}

A moment's reflection will show that a number of interpretations of 'P' and 'R' will make each member of (□1) true, but that no interpretation will render true each member of (□2). That is, the criteria to be met by any pair of concepts or sets which are to interpret 'P' and 'R', respectively, in such a way as to make each member of (□1) express a truth are readily satisfiable, while the criteria to be met by interpretations which would make (□2) true are unsatisfiable. Instead of "criteria", we can speak of the complex *properties* defined by the sets (□1) and (□2). In particular, the property to be met by any pair <□, □> interpreting 'P' and 'R' respectively in (□1) is that

(P1) □ is a non-empty unary relation; □ is a binary relation whose domain includes □; □ is not included in the range of □.

while the property established by (□2) is that

(P2) □ is a non-empty unary relation; □ is a binary relation whose domain includes □; □ is not included in the domain of □.

If there are n non-logical terms in a set of sentences, the property defined by the set will be a complex n-place property. As Hilbert sees it, each set of axioms defines such a property, or "scaffolding of concepts." The nonlogical terms which play the schematic role of 'P' and 'R' above are, in Hilbert's sentences, the geometrical 'point', 'line', 'between', etc.

Say that the complex property defined in this way by a set of partially-interpreted sentences is *consistent* just in case it is not self-contradictory, i.e. just in case some series of concepts or sets could have that property.[9] (P1) is clearly consistent in this sense, while (P2) is not. Though Hilbert does not speak explicitly of property-consistency, his proofs immediately show consistency in this sense. For the interpretation used to prove the consistency of a set □ is a series of concepts or sets satisfying the property defined by □ (assuming, again, the consistency of the background theory). The parallel notion of independence is straightforward: the property defined by □ is independent of that defined by □ just in case the property defined by □;~□ is consistent. This, again, is immediately demonstrated by Hilbert's consistency-proof.

Where □ is a set of sentences, call a sequence of sets or concepts satisfying the property defined by □ a *□-structure*. The existence of a □-structure implies the consistency of the property defined by □. This in turn implies the syntactic consistency of □ itself. When □ is a set of first-order sentences, the syntactic consistency of □ implies the existence of a □-structure,[10] so that in the first-order case (with which Hilbert is primarily concerned), we come full circle; syntactic consistency and property-consistency are in this case equivalent.

II. Frege

II.1

Frege was clearly of the view that principles of logic can be expressed as syntactic transformation-rules; his own formal system used precisely the kinds of topic-neutral, syntactically-specified rules just discussed. Given a set of sentences in Frege's *Begriffsschrift* notation, a Hilbert-style consistency proof will demonstrate that no contradiction is derivable from that set via Frege's own rules of logic.[11] Hilbert-style independence-proofs, similarly, would successfully demonstrate the non-derivability in Frege's system of one of Frege's sentences from a set thereof. It looks, therefore, as if Frege's assessment of Hilbert's demonstrations as "failures" is simply wrong.[12]

But this conclusion would be hasty. Consistency and independence, along with all of the fundamental logical relations, hold as Frege sees it not between sentences, but between the nonlinguistic propositions or *thoughts* expressed by sentences.[13]

> When one uses the phrase 'prove a proposition' in mathematics, then by the word 'proposition' one clearly means not a sequence of words or a group of signs, but a thought; something of which one can say that it is true. And similarly, when one is talking about the independence of propositions or axioms, this, too, will be understood as being about the independence of thoughts.
> ...
> We have to distinguish between the external, audible, or visible which is supposed to express a thought, and the thought itself. It seems to me that the usage prevalent in logic, according to which only the former is called a sentence, is preferable. Accordingly, we simply cannot say that one sentence is independent of other sentences; for after all, no one wants to predicate this independence of what is audible or visible.[14]

The thoughts, which strictly speaking are the items proven, and the sentences which occur in formal deductions, are connected as Frege sees it by the fact that each sentence in a deduction expresses a determinate thought. The sentential inference-rules are designed so that a sentence follows from previous sentences via a rule only if the thought expressed by that sentence follows logically from the thoughts expressed by the previous sentences.

To keep these distinctions clear, let us use the term 'derivation' for a series of *sentences* countenanced in the usual way by a formal system, and the word 'proof' for a series of *thoughts* each of which is either taken as premise or is a logical consequence of previous members of the series. Because the sentences of a derivation all express determinate thoughts and the inference-rules mirror logical relations, to derive a sentence in a system like Frege's is to prove the thought it expresses. If a sentence s is derivable from a set ▯ of sentences, then the thought expressed by s is provable from, and hence a logical consequence of, the thoughts expressed by the members of □.

Though each sentence expresses only one thought, a given thought is expressible by distinct sentences. The most striking examples of the different expressions of a thought occur when sentences of dramatically different syntactic structure express the same thought. In such cases, the distinct sentences each reflect a more or less thorough "decomposition" of logically-complex concepts or relations into their constituents. For example, while the thought

(1) *Every cardinal number has a successor* [15]

can be expressed by (Frege's version of)

(a) (□x)(Nx (□y) Syx),

the result of the conceptual analysis of *cardinal number*, given by Frege in the *Grundlagen,* is that the thought can also be expressed by the sentence

(b) (□x)((□F)(x = ext(□ ~ F)) (□y) Syx)[16].

Further analysis of the relation of equinumerosity (~) and of successor (S) shows that the same thought is expressible by a yet more-complex sentence.[17]

That the same thought is expressed by each of these sentences is central to Frege's picture of conceptual analysis and proof: in deriving either (a) or (b) from a given set of sentences, one shows that the thought (1) is a logical consequence of the thoughts expressed by that set.[18] The hope of Frege's logicist project is that analyses of the central arithmetical concepts will show each truth of arithmetic to be expressible by a sentence which can be derived using only principles of logic. As Frege puts it, when trying to prove the truths of arithmetic from simpler propositions,

> ... we very soon come to propositions which cannot be proved so long as we do not succeed in analysing concepts which occur in them into simpler concepts or in reducing them to something of greater generality. Now here it is above all cardinal number which has to be either defined or recognized as indefinable. This is the point which [the *Grundlagen*] is meant to settle. On the outcome of this task will depend the decision as to the nature of the laws of arithmetic.[19]

The point worth stressing for present purposes is that a thought expressible by a relatively simple sentence can be seen, as a result of conceptual analysis, to be expressible by each of a series of syntactically more-complex sentences. In the series of increasingly-complex sentences expressing a given thought, each member is derivable from sentences from which its less-complex predecessors are not. As Frege puts it,

> In the development of a science it can indeed happen that one has used a word, a sign, an expression, over a long period under the impression that its sense is simple until one succeeds in analysing it into simpler logical constituents. By means of such an analysis, we may hope to reduce the number of axioms; for it may not be possible to prove a truth containing a complex constituent so long as that constituent remains unanalysed; but it may be possible, given an analysis, to prove it from truths in which the elements of the analysis occur.[20]

This is the sense in which conceptual analyses and the definitions which express them give us insight, as Frege puts it, "into the logical linkage of truths."[21]

Because each thought is generally expressible by a number of different sentences, there is much more to the relation of *provability* than is evidenced by the relation of *derivability*. Where p is a thought and s a sentence expressing it, □ a set of thoughts and □ a set of sentences expressing □: the derivability of s from □ guarantees that p is a consequence of □, but the fact that s is not derivable from □ is no guarantee that p is not a consequence of □. For s's non-derivability from □ is entirely compatible with the existence of some s' and □' expressing p and □ respectively, such that s' is derivable from □'. Relations between sentences of a well-designed formal system can always, for Frege, provide a positive test of provability and consequence, but never a negative test.

II.2

For Frege, a set of thoughts is consistent just in case no contradiction is provable from it. Given the above, then, if a contradiction is derivable in a standard syntactic system from a set □ of sentences, then the set of thoughts expressed by the members of □ is inconsistent in Frege's sense. But the converse is false: the fact that no such contradiction is derivable from □ is no guarantee that the set of thoughts expressed is, in Frege's sense, consistent. In short: syntactic inconsistency guarantees Frege-inconsistency, but syntactic consistency is no guarantee of Frege-consistency. Similarly, syntactic independence is no guarantee of independence in Frege's sense, since p is independent of □, for Frege, just in case p is not *provable* from □.

The inconsistency of a set of thoughts despite the syntactic consistency of a set of sentences expressing those thoughts is not simply a far-fetched possibility for Frege. It is of the very essence of his views on analysis and thoughts that such cases are to be expected. Consider for example the Peano axioms expressed in their usual form, i.e. with simple predicate letters for the concept *natural number* and the relation *successor*. If Frege's logicist thesis is correct, these sentences express truths of logic, so that any set containing the negation of one of them is, in Frege's sense, inconsistent. That is, any such set expresses an inconsistent set of thoughts.[22] But such a set of sentences will not, of course, generally be syntactically inconsistent. It is a trivial matter to interpret the negations of Peano-axiom sentences in such a way that they express truths, and hence to give a Hilbert-style demonstration of their syntactic consistency. But this rather mundane fact is no argument against logicism. It is because logical analysis can reveal previously-unrecognized relations of logical consequence that such analysis is central to the logicist project, and it is just this fact which distinguishes Hilbert's sentential relations from Frege's logical relations.

In a review of Peano, Frege stresses the role of analysis in his own conception of formal logic and foundational work:

> What I am aiming for, then, is uninterrupted rigour of demonstration and maximal logical precision, together with perspicuity and brevity.
>
> I cannot so definitely specify what aim it is that Peano is pursuing with his conceptual notation or mathematical logic: I have to rely largely upon conjecture. ... This much I think I can gather from it, that an examination of the foundation of mathematics is not what initiated it - nor has it been a determinant for its mode of execution. For straightaway, in §2 of [Peano's work[23]], brief tags are introduced for the classes of the real numbers, the rational numbers, the prime numbers, etc., which means that all these concepts are assumed as already familiar. The same thing happens with the meanings of the operation-signs '+', '-', 'x', ' ', etc., from which it is to be gathered that an analysis of these logical structures into their simple components was not the intention. And since, without such an analysis, an investigation like the one I projected is impossible, such an investigation could not have been among Mr. Peano's intentions.[24]

II.3

Just as the consistency and independence of axioms concerning cardinal numbers and successor will turn on the analysis of the concepts *cardinal number* and *successor*, so too presumably the consistency of thoughts concerning points and lines will turn on the analysis of the concepts *point*, *line*, and *between*, among others. Frege reports to Hilbert that in his own unfinished investigations into the foundations of geometry, he thought he could "make do with fewer primitive terms."[25] Presumably, this indicates that as far as Frege is concerned, some of the concepts expressed by these terms can be analysed in terms of others.

Since none of Frege's axiomatic investigations into geometry have survived, we can only guess at the kinds of analyses he would give. But the results of any such reduction would be the same: they would show that some of the sentences which are Hilbert-independent of one another express thoughts which are not Frege-independent. If, for instance, it turned out on careful analysis of the geometrical concepts that the thought

Point B lies between points A and C

logically implies the thought

Point B lies between points C and A

then Hilbert's axiom

(II.1) If A, B, C are points of a straight line and B lies between A and C, then B lies between C and A

would be, from Frege's point of view, superfluous. The thought expressed by II.1 would not be independent of the other axioms of order, despite the fact that the sentence II.1 is clearly independent, in Hilbert's sense, from those axiom-sentences.

Similarly, if the analysis of the geometrical concepts alone suffices to establish e.g. that the congruence of two triangles follows from the congruence of two sides and the enclosed angle of those triangles, then no set of sentences containing the negation of this axiom would express a set of Frege-consistent thoughts. This despite the fact that such a set will in general be Hilbert-consistent.

Frege does hold that a number of the axioms of geometry are independent of one another, and hence that no amount of conceptual analysis will show them to be reducible to truths of logic or provable from one another. There is no evidence that he takes any of Hilbert's specific independence-claims about geometrical axioms to be false. Frege's objection is rather that nothing Hilbert demonstrates in the *Foundations of Geometry* can support such claims.

In general, the syntactic consistency of a set of sentences is insufficient to guarantee what Frege would consider the consistency of a set of thoughts expressed by those sentences. The case is the same with respect to property-consistency and the existence of □-structures. The consistency of the property defined by a set □ of schematic sentences, and the existence of a □-structure, all turn on the syntactic form of the sentences in □. For Frege on the other hand, two sets of sentences indistinguishable as to syntactic form can express sets of thoughts one of which is consistent while the other is not. Again, the results demonstrated by Hilbert's proofs do not imply the results with which Frege is concerned, and on which, as Frege sees it, the foundations of mathematics is based.

III. Geometrical Concepts

In addition to the consistency of thoughts, Frege recognizes a kind of consistency which applies to properties or, as he would say, to concepts. Concepts, for Frege, come in levels. A first-level concept is one under which objects fall; thus the ordinary concept *point*, under which fall all and only points, is first-level. A second-level concept is one under which first-level concepts fall; thus *being non-empty* is a second-level concept. A concept can be expressed by an open sentence; for example, 'x is a point' refers to the concept *point*, that concept under which fall all and only points.

The sense in which concepts can be consistent is never made entirely clear by Frege, though two principles to which he is clearly committed are

(1) that the only way to show the consistency of a concept or collection of concepts is to exhibit some thing that falls under it (or them);[26]

and

(2) that the consistency of a collection of concepts does not guarantee the existence of something falling under them; similarly for individual concepts.[27]

Regarding this last point, Frege's often-repeated example is that the joint consistency of the three concepts

x is intelligent; x is omniscient; x is omnipotent

does not guarantee the existence of an intelligent, omniscient, omnipotent being. Though Frege does not explicitly connect the consistency of concepts with that of thoughts, his practice indicates a conception of consistency on which the consistency of a finite collection of concepts is tantamount to the consistency of the existentially-quantified thought that something falls under them.

The question, now, is whether Hilbert's consistency-proofs demonstrate anything which Frege ought to have recognized as the consistency of concepts (or relations). Before turning to this question, though, we pause to consider Frege's claim (1) above. Michael Resnik argues that Frege is mistaken here:

> Consider the [concepts defined by the] open sentences
> x is horse-like
> x has a single horn.
> Let us suppose that they are simple. Then, contrary to Frege, we need not show that unicorns exist in order to demonstrate their joint consistency. Instead it suffices to show that the quantificational schema "Fx.Gx" is consistent, and that can be done by assigning new interpretations to 'F' and 'G' and 'x.' As this is the very method implicit in Hilbert's consistency proof for geometry, one can understand why he did not reply to Frege's objection that his method could not be applied.[28]

The Hilbert-style proof does indeed show the syntactic consistency of the pair of open sentences. But it is not clear why the proof should be taken to show the consistency of the Fregean concepts. If the consistency of the pair of concepts is equivalent to the consistency of the thought

(U) *There is something that is horse-like and has a single horn*,

then the consistency-claim requires it to be established that no further analysis of the concepts *horse-like* and *has a single horn* will reveal a contradiction in (U); and this will not be established by a re-interpretation of the terms. It would, of course, be established by producing a unicorn.

Perhaps it is just such further analysis which Resnik means to rule out via the supposition that the sentences are "simple." But if so, this would seem to undermine the objection. For the reason that Hilbert-style consistency-proofs fail to show Frege-consistency is precisely that one cannot in general make such a supposition. Consider for example the pair of first-level concepts expressed by the open-sentence pair:

(V) x is a natural number; x has no successor

A Hilbert-style consistency-proof can easily be given for this pair of open sentences. But it is clear that in Frege's view the concepts expressed in (V) are inconsistent.

As we have seen, Hilbert's consistency-proofs demonstrate the consistency of the property defined by a set of sentences when the non-logical terms are read as purely schematic. They do so in a way which corresponds to Frege's own views about such consistency-demonstrations, since they exhibit something (in this case, a □-structure of first-level arithmetical concepts and relations) which satisfies the defined property. As Frege would put it, Hilbert's consistency-proofs demonstrate the consistency of a complex second-level relation and hence of its defining conditions.[29]

But the consistency of this relation has little to do with the consistency of any first-level geometrical concepts or thoughts. For the consistency of the relation defined by a set □ of schematic sentences is just a matter of the existence of *some* interpretation under which the members of □ express truths. The subject-matter of that interpretation is, for Hilbert's purposes, irrelevant:

> ... it is surely obvious that every theory is only a scaffolding or schema of concepts together with their necessary relations to one another, and that the basic elements can be thought of in any way one likes. If in speaking of my points I think of some system of things, e.g. the system: love, law, chimney-sweep ... and then assume all my axioms as relations between these things, then my propositions, e.g. Pythagoras' theorem, are also valid for these things. In other words: any theory can always be applied to infinitely many systems of basic elements.[30]

The consistency of the Fregean thoughts, on the other hand, has to do essentially with the thoughts expressed under the geometrical interpretation of those sentences. As Frege puts it, Hilbert

> widen[s] the area so that Euclidean geometry appears as a special case; and in this wider area he can now show lack of contradiction by examples; but only in this wider area; for from lack of contradiction in a more comprehensive area we cannot infer lack of contradiction in a narrower area; for contradictions might enter in just because of the restriction. The converse inference is of course permissible. Hilbert was apparently deceived by the wording. If an axiom is worded in the same way, it is very easy to believe that it is the same axiom. But it depends on the sense; and this is different, depending on whether the words 'point', 'line', etc. are understood in the sense of Euclidean geometry or in a wider sense.[31]

For Hilbert, the logical relations between geometrical axioms are independent of the particular subject-matter of those axioms, with the result that satisfaction by constructions out of the real numbers (or chimney-sweeps) gives as much information about these logical relations as would satisfaction by geometrical concepts. For Frege on the other hand, since logical relations between axioms depend on relations between the concepts and objects mentioned in those axioms, geometrical axioms can never be shown consistent by demonstrating that non-geometrical concepts satisfy Hilbert's defined conditions.

IV. Frege and Model Theory

Contemporary models differ in two respects from the interpretations Hilbert uses in the *Foundations of Geometry*. First, models today are typically constructed out of "pure" sets, and the background theory is an axiomatic set theory, rather than a theory of the real numbers. Secondly, a modern model incorporates a domain for unrestricted quantifiers; this is generally thought of as a "completed infinite totality" (whatever precisely this means) in a way which probably would not have been congenial to Hilbert's views of mathematical existence. But the overall approach to consistency-proofs is the same. By giving a model of a set of sentences, one shows (a) that the complex property defined by the set is non-empty, which implies (b) that the set is syntactically consistent. The model-theoretic consistency of a set of sentences is equivalent to (a), which as we have seen does not guarantee the Fregean consistency of any set of thoughts expressed by those sentences. The model-theoretic independence of a sentence from a set of sentences, for the same reasons, fails to show the Frege-independence of the thoughts expressed.

Frege's conception of logic differs from the model-theoretic conception not just because interpretations, or models, make no sense with respect to thoughts, but also because the claims established by the use of models have no relevance to what Frege would characterize as logical results. It is of the essence of the model-theoretic approach that the content of non-logical terms is irrelevant; sentences with the same syntactic form have all the same model-theoretic properties. It is essential to Frege's view on the other hand that the content of the non-logical terms is pivotal, and that sentences with the same syntactic form will often express thoughts with dramatically different logical properties.

Despite these differences, Michael Dummett has urged a view of Frege on which the model-theoretic concepts and distinctions, though not explicit in Frege's writings, are there implicitly. As Dummett sees it, Frege's views about the consistency of theories can be expressed in modern model-theoretic terminology, and can thereby be seen to be in conflict with modern results. On Dummett's account, Frege's views can be summarized in the following four propositions:

(i) The consistency of a theory requires proof;
(ii) the consistency of a theory does not imply the existence of a model for it;
(iii) anyway, the only way of proving the consistency of a theory is to provide a model for it; and
(iv) even if it were possible to prove the consistency of a theory by some other means, what matters is the existence of a model and not the bare formal consistency.[32]

The first of these propositions is terminologically odd: taking a theory in Frege's sense as a collection of thoughts and their consequences, (i) is false. For, as Dummett notes, Frege holds (rather oddly) that axioms are always true, and hence that there is no need to demonstrate the consistency of axioms and their consequences.[33] Dummett's intended reading, however, is the unproblematic

(i') It requires proof that a concept is not self-contradictory.[34]

The terminological issues are more problematic with respect to (ii) and (iii). Because a theory for Frege is not the kind of thing which can intelligibly be said to have a model, neither of these claims, read literally, is endorsed by Frege. Dummett's further characterization is as follows:

> Proposition (ii) is a rejection of that variety of formalism according to which the existence of mathematical entities is tantamount to the consistency of the theory relating to them. It is in fact incorrect for a first-order theory, in view of the completeness theorem, but true for a higher-order theory;[35]

and the later gloss of (ii) is

(ii') One cannot infer, from the fact that a concept is not self-contradictory, that an object falls under it[36]

As noted above, Frege clearly holds (ii'); recall his theological example. But this is not the view, reasonably expressed as (ii), which is shown by the completeness theorem to be incorrect. The completeness theorem shows that some concepts do have things falling under them: specifically, that any relation definable by a syntactically-consistent set of first-order sentences has a set-theoretic structure falling under it. This would probably have been a surprise to Frege. But it is no refutation of the Fregean claim (ii') - which, incidentally, would seem to be true.

Dummett claims that "Proposition (iii) is not only false, but could have been known by Frege to be false." As evidence of the falsehood of (iii), Dummett points out that an infinitely-axiomatized theory can be shown consistent by showing that each finite subset has a model, without providing a model of the whole. And as a restatement of the Fregean view characterized by (iii), Dummett offers:

(iii') the only way to establish that a concept is not self-contradictory is to prove that an object falls under it.[37]

The method of giving models of each finite subset of a set of sentences shows the syntactic consistency of that set. Thus the view which Dummett has shown to be false is (iii), with 'consistency' read as 'syntactic consistency.' But this does not seem to have been a view held by Frege, who in any case had little interest in syntactic consistency. Frege did, as Dummett demonstrates, hold (iii'). The distinction between Frege's view and (iii) can perhaps best be seen by noting that while (iii) is as Dummett says clearly false, it is not at all clear that there is a means of demonstrating the consistency of concepts other than the one Frege gives: namely, of exhibiting an appropriate object.

Lastly, it is important to point out that Frege's view of semantics is not a model-theoretic view. According to Dummett,

> Much the philosophically most important of these four contentions, however, is the last one (iv), the thesis, namely, that, even if the consistency of a theory could be proved without establishing that it had a model, what matters is that the theory should have a model, and not merely that it be consistent. Frege's reason for holding this is his general theory of meaning...[38]

Because the sense of a sentence is given by associating with it determinate truth-conditions which in turn require an association between words and referents, Dummett continues,

> Any mathematical theory, therefore, if it is to have a specific interpretation which renders the sentences of the theory statements with a determinate meaning, must be related to a definite model: the variables of the theory must be assigned a definite domain over which they are taken to range, and the primitive individual constants, predicate-symbols, function-symbols, etc., of the theory given referents (of suitable types) in or over that domain. To establish a theory as formally consistent is not, in itself, to confer a meaning on it, or to explain the meaning which it has: that can be done only by describing a model for the theory.[39]

If Dummett means simply that for Frege it is important that the words and sentences used in mathematics express senses and refer to determinate referents, then there is little here to take issue with. But it is misleading to express this view in terms of models. For to give a *model* of a sentence is to give an interpretation on which that sentence is true. And there is nothing in

Frege which indicates that the meaningfulness of a false sentence turns on the existence of some *re*interpretation on which the sentence expresses a truth. As Dummett says, to understand a sentence is in part to understand the conditions under which it would be true; but this is an understanding of the conditions under which the thought actually expressed would be true, not an understanding of what *other* actually-true thought the sentence might express.

Dummett's claim, then, that Frege's "procedure is exactly the same as the modern semantic treatment of the language of predicate logic,"[40] and that the "modern distinction between the semantic (model-theoretic) and the syntactic (proof-theoretic) treatments of the notion of logical consequence ... is implicit in his writing,"[41] cannot be right. Frege's procedure is certainly *semantic*, in that it is concerned with the meanings, and not simply with the syntax, of sentences. But it is not semantic in anything like the modern model-theoretic sense, in which the emphasis is on reinterpretations of non-logical terminology. Similarly, Frege clearly makes a distinction between two conceptions of consequence - what we have called the relations of *provability* and *derivability*. But though the latter is our modern syntactic, proof-theoretic notion, the former is not the relation of model-theoretic consequence.[42]

V. Ultimate Analyses?

The gulf between syntactic consistency and Frege-consistency is most vivid with respect to syntactically-consistent sets of sentences which express sets of thoughts also expressible by syntactically-inconsistent sets of sentences. This raises the question whether in Frege's view there is a "final" level of analysis at which consistency and syntactic consistency coincide. Is there a system of maximally fine-grained linguistic representations of thoughts such that each inconsistent set of thoughts is represented in this system by a syntactically-inconsistent set of sentences? Frege, as far as I can tell, simply does not consider this question, and nothing he says appears to commit him to either an affirmative or a negative answer. Nevertheless, perhaps a few remarks are in order.

In order to demonstrate positive consequence-results, i.e. claims of the form "p is a logical consequence of □", for p a thought and □ a set of thoughts, Frege requires p and □ to be expressed by a sentence s and set □ of sentences, respectively, such that s is derivable from □. This can, and in the case of the logicist project, does, require a certain amount of detailed analysis, since the most familiar sentences used to express p and the members of □ may not be derivable one from another. In short, the demonstration of positive consequence-results will often require conceptual analysis, of the kind Frege takes himself to be providing in the *Grundlagen.* What such demonstrations do not require, however, is the claim that the analysis given is in any sense final; there is no need to show, and indeed Frege never attempts to show, that further, more-detailed analysis is impossible. Thus the demonstration of logicism requires no particular view about the existence of a final analysis of thoughts, or of a language in which all logical relations are reflected syntactically.

One reason to suppose that Frege would have denied the existence of such an ultimate level of analysis stems from his remarks about the "re-carvability" of thoughts.[43] Frege holds, at least in the *Grundlagen*, that a given thought can be expressed by two sentences neither of which is obtainable from the other by analysis of the simple components of either. Thus, for example, the same thought can be expressed via sentences of the form "R(a,b)" and "f(a)=f(b)". Because neither sentence is obtained from the other by a mere expansion of atomic parts, there is no obvious sense in which one expresses a fuller analysis of the thought than does the other. Frege's acknowledgment of such pairs is a reason to give a negative answer to the question of this section. The reason is inconclusive, however, since Frege might well have held, if pressed, that such incomparable pairs of sentences were always further reducible to some common sentence expressing a final analysis.

In the end, the answer to this question will have little impact on the assessment of Hilbert's consistency-proofs. If a set □ of sentences gives the final analysis of a set □ of thoughts, then □ is syntactically consistent just in case □ is Frege-consistent. But this will not mean that a Hilbert-style consistency-proof will demonstrate Frege-consistency. For in order to demonstrate the consistency of □, one will need not just the Hilbert-style proof of syntactic consistency, but also a proof of the auxiliary claim that □ does indeed represent the "ultimate" analysis of □. One will need a demonstration that the analysis is finished. And this, of course, is a claim which turns not just on the syntactic form of the sentences in □, but on the content of their non-logical terms. It is the kind of claim, in short, which is indemonstrable via a Hilbert-style consistency proof.

VI. Conclusion

Frege's assessment of Hilbert's work as "on the whole a failure" is not just badly wrong, but terribly short-sighted. Given the growing importance at the time of Frege's own work of the kind of formal investigations undertaken by Hilbert, Frege's steadfast refusal to acknowledge the existence of a different but nonetheless extremely valuable conception also called (perhaps unfortunately) 'consistency' is striking. As the 19th-century revolution in geometrical methods and results had made clear, the view of axioms as only partially-interpreted, and hence as capable of characterizing similarities across abstract structures, was an undeniably fruitful one. Further, Frege's own interest in formal systems and in relations of syntactic derivability ought to have enabled him to see the importance of Hilbert's method for demonstrating non-derivability results, even if these did not show what he would have called 'independence.'

Nevertheless, Frege was right to hold that the method of interpretations holds no promise for demonstrating what he took to be consistency and independence results. The point here is not merely terminological: Frege takes it that the consistency of a set of thoughts, and the independence of thoughts from one another, have important philosophical ramifications, none of which are demonstrable via Hilbert's method.

For Frege, the logical relations are epistemologically significant: If a thought □ is a logical consequence of a set □ of thoughts, then knowledge of the members of □ can provide grounds for the knowledge of □. The independence of □ from □ shows that some source of knowledge other than that given by □ is required to ground □. The distinction between analytic and synthetic, which Frege takes to be fundamentally epistemological, turns on the logical relations as well: a thought is analytic if it is provable from the laws of logic alone, and is synthetic if it is independent of those laws.[44] And finally, the logical relations are of ontological significance: the provability of □ from □ shows that □'s ontological commitments are no greater than □'s, while the independence of "there are □'s" from □ guarantees that □ carries no commitment to the existence of □'s.

None of these results is demonstrable via Hilbert's method. The fact that knowledge of the members of □ suffices to ground knowledge of □ is entirely compatible with the *in*dependence, in Hilbert's sense, of the sentence expressing □ from those expressing the members of □. This will be the case whenever, as in the case of arithmetic if Frege is right, the epistemological dependence turns on analysis of the constituent concepts of the thoughts in question. In just such situations, thoughts which Frege takes to be analytic will be expressible by sentences which are Hilbert-independent of the principles of logic, so that Hilbert-independence from logic is no guarantee of Fregean syntheticity. And finally, for the same reasons, the independence in Hilbert's sense of a sentence asserting the existence of □'s from a set of sentences tells us nothing about the ontological commitments of that set.

---

[1]Versions of this paper were delivered at the University of Notre Dame, the University of California/Riverside, and the 1995 meeting of the Society for Exact Philosophy. Thanks to members of each audience for helpful comments, and especially to Mic Detlefsen and Aldo Antonelli for comments and criticism. Thanks also to the Center for the Study of Language and Information at Stanford University for its hospitality, and to the University of Notre Dame for a research leave.

[2]See e.g. *Die Grundlagen der Arithmetik*, trans. as *The Foundations of Arithmetic*, J.L. Austin, ed. (Evanston, Ill.: Northwestern, 1978) §14.

[3]At times, Frege's view seems to be that such demonstrations are in principle impossible; see e.g. the letter to Liebmann of 29.7.1900 (*Philosophical and Mathematical Correspondence*, Gabriel, et al eds. (Oxford: Blackwell, 1980) (hereafter, PMC) pp 90-91; also "On Formal Theories of Arithmetic," *Collected Papers on Mathematics, Logic and Philosophy*, McGuinness ed. (Oxford: Blackwell, 1984) (hereafter, CP) pp 112-121, esp 119-20. At other times the view seems rather to be that no currently-available method is successful. Frege does investigate a potential method for demonstrating consistency in "On the Foundations of Geometry: Second Series," (CP 293-340 esp 337 ff.) but never follows through on this idea.

The skepticism about demonstrability in principle appears again in Frege's 1910 notes to Jourdain's "Gottlob Frege," reprinted in PMC 179-206, esp 183.

[4]Hilbert, *Foundations of Geometry*, L. Unger, ed. (La Salle, Ill: Open Court, 1971), translation of the 10th edn. of *Grundlegung der Geometrie*. (Hereafter, Hilbert.)

[5]See e.g. letter to Liebmann of 29.7.1900, PMC 90-91.

[6]One actually needs the further constraint that not both of u and v are 0. The *ratio* [u:v:w] is the equivalence-class of triples <u',v',w'> such that u:v = u':v' and v:w = v':w'. The interpretation outlined here is part of that given by Hilbert in §9; see Hilbert for a specification of □.

[7]Hilbert, §9 (p 30).

[8]To clarify: Hilbert has not at this point specified a syntactic deductive system, and does not view logical deduction as formal symbol-manipulation. He does however view logical deduction as independent of the meanings of the non-logical (here, the geometrical) terms, which makes his implicit principles of deduction syntactically specifi*able*, though not explicitly so specified (or specified at all, for that matter).

[9]In order to be precise here, one would have to say more about the sense of "could" intended in the last phrase. Nothing will turn on such precision, though; claims in what follows about property-consistency will hold on any reasonable understanding of this "could."

[10]This is Gödel's completeness theorem for first-order logic.

[11]Excluding the *Grundgesetze*'s Axiom V.

[12]See e.g. Frege's letter to Liebmann of 29.7.1900: "Clear and inventive as [Hilbert's *Foundations*] is in many points, I think that it is on the whole a failure and in any case that it can be used only after thorough criticism." PMC 90.

[13]A *thought* (*Gedanke*) is an objective, non-mental entity; it is the primary bearer of truth and falsehood, and is what a declarative sentence expresses. The view of nonlinguistic items as bearers of the logical relations antedates the sense-reference distinction and

the terminology of thoughts; see e.g. *Begriffsschrift* in *From Frege to Gödel*, van Heijenoort, ed. (Cambridge: Harvard, 1967) 5-82 in which the logical relations hold between nonlinguistic *contents*.

[14]"On the Foundations of Geometry: Second Series" 332. I have altered Kluge's translation to render "Satz" as "sentence."

[15]I will use italicized English sentences in order to indicate that it is the thought expressed, not the sentence, which is under discussion.

[16]Here, "~" is short for Frege's relation of equinumerosity; see *Grundlagen* §72.

[17]See *Grundlagen* §§72, 76.

[18]I will for brevity speak of the set of propositions *expressed by* a set of sentences when what I mean is the set of propositions expressed by the members of that set of sentences.

[19]*Grundlagen* §4.

[20]From the posthumously-published "Logic in Mathematics," dated by the editors at 1914, in *Posthumous Writings*, Hermes et al eds. (Oxford: Blackwell, 1979) 209.

[21]"On the Foundations of Geometry: Second Series" 302.

[22]See *Grundlagen* §14 and "On Formal Theories of Arithmetic" 112.

[23]*Notations de Logique Mathematique: Introduction au Formulaire de mathematiques* (Turin: Guadagnini, 1894).

[24]"On Mr. Peano's Conceptual Notation and My Own" CP 234-248, esp 237.

[25]Letter to Hilbert of 27.12.1899, PMC 34.

[26]See e.g. letter to Hilbert of 6.1.1900, PMC 43; *Grundlagen* §§95, 109; "On Formal Theories of Arithmetic" 120; *Grundgesetze* Vol II §144 (portions translated in *Translations from the Philosophical Writings of Gottlob Frege*, Geach and Black eds (Oxford: Blackwell, 1980).)

[27]See e.g. *Grundlagen* §94; letter to Hilbert of 6.1.1900, PMC 47.

[28]Michael Resnik, "The Frege-Hilbert Controversy," *Philosophy and Phenomenological Research* 34 (1973/4) pp 386-403. Quotation on p 397. A similar criticism of Frege is offered by Gregory Currie; see his *Frege: An Introduction to His Philosophy* (Sussex: Harvester, 1982) pp 78-9.

[29]Frege refers to these defining conditions or the schematic sentences that express them as "pseudo-axioms;" see e.g. "On the Foundations of Geometry, Second Series" 332 ff.

[30]Letter to Frege of 29.12.1899, excerpt by Frege; ellipsis Hilbert's or Frege's. PMC 40-41

[31]Letter to Liebmann 29.7.1900, PMC 91

[32]"Frege on the Consistency of Mathematical Theories" (FCMT) in *Studien Zu Frege*, M. Schirn, ed. (Stuttgart/Bad Constalt: Fromann-Holzboog, 1975) pp 229-242. Quotation p. 229.

[33]See e.g. letter to Hilbert of 27.12.1899, PMC 37: "From the truth of the axioms it follows that they do not contradict one another. There is therefore no need for a further proof." Also "On The Foundations of Geometry: First Series," CP 275.

[34]FCMT 234

[35]FCMT 229

[36]FCMT 234

[37]FCMT 234

[38]FCMT 231

[39]FCMT 232

[40]*Frege: Philosophy of Language* (Cambridge: Harvard, 1981) (hereafter, FPL) p. 90.

[41]FPL 81

[42]Hans Sluga has criticized Dummett's attribution of model-theoretic notions to Frege; see his *Gottlob Frege* (London: Routledge & Kegan Paul, 1980) pp 180-181.

[43]See e.g. *Grundlagen* §64.

[44]Frege's characterization of the analytic/synthetic distinction at *Grundlagen* §3 is that the analytic truths are those provable from laws of logic and definitions. Because definitions are stipulative, they are always eliminable, which is to say that any *thought* provable via laws of logic and definitions is provable via laws of logic alone. See e.g. *Begriffsschrift* §24, *Grundgesetze* §2, and my "Frege's Reduction," *History and Philosophy of Logic* 15 (1994), 85-103. Thus as Frege says at e.g. *Grundlagen* §87, the logicist reduction was to have shown that the truths of arithmetic were analytic.