My Take on the Frege-Hilbert Controversy

Michael Barkasi
mtb2@rice.edu

Sept 2011

1 Introduction

What follows is a mildly polished write up of the material given in a casual presentation (6/2/2011) on the "Frege-Hilbert debate" to STUMPS (Student Mathematical Philosophy Seminar) at Rice. One will find a list for further reading at the end. I implore the reader for charity and ask that she understand that the following is not intended to be groundbreaking, original, or a substantial piece of academic research. (My treatment is largely drawn from the sources in the reference section.) As the title says, it is a quick write up of my take on an interesting historical debate.

2 Frege's Logicism

Frege sets out and executes his logicist project in three seminal works: *Begriffsshrift* (1879), *Grundlagen der Arithmetik* (1884) and *Grundgesetze der Arithmetik* (1893/1903). In *Grundlagen* he states his goal as:

> Now here it is above all Number which has to be either defined or recognized as indefinable. This is the point which the present work is meant to settle.

After giving the desired definitions he sketches out how basic facts of arithmetic follow from them and "principles of logic" alone, while in *Grundgesetze* he carries out a more rigorous derivation of the Dedekind-Peano axioms from the definitions and axioms for his system of (second-order) logic plus his theory of extensions.

But what does Frege mean when he says he wants to *define* number? By the time of writing Dedekind's axioms for the natural numbers (and successor function) were well known. They are, more or less, as follows:

Dedekind-Peano Axioms:

1. Zero is a natural number.
2. The successor of every natural number exists and is a natural number.
3. Zero is not the successor of any natural number.
4. The successor function is injective.
5. If a subset of natural numbers contains zero and is closed under the successor function, then that subset is all of the natural numbers.

Letting 'Nx' intuitively stand for "x is a natural number" and '$s(x)$' be the successor of x, these can be written:

Dedekind-Peano Axioms, Formalized:

1. $N0$
2. $\forall x(Nx \Rightarrow \exists y(Ny \wedge s(x) = y))$
3. $\neg\exists x(Nx \wedge s(x) = 0)$
4. $\forall x \forall y(s(x) = s(y) \Rightarrow x = y)$
5. $\forall X((X0 \wedge \forall x(Xx \Rightarrow Xs(x))) \Rightarrow \forall y Xy)$

The goal is to give definitions for the relation N, the constant 0 and the function s so that the above formal sentences can be derived (once 'Nx,' '0,' and '$s(x)$' have been replaced in them by the definitions) from *just* "principles of logic." Setting aside his defunct theory of extensions, the logic which Frege actually uses is equivalent to the standard axioms and rules of inference for second-order logic *plus* what's come to be known as *Hume's Principle*. (His theory of extensions being defunct, of course, because it engenders Russell's Paradox.)

We begin by working in a second-order language $\mathcal{L}$ whose signature contains only the function symbol #, which denotes a function from unary relations to objects and is intuitively read "the number of … ". (See Appendix A for a complete definition of $\mathcal{L}$.) Next we define a deduction system for $\mathcal{L}$ which consists of the following axioms and rules of inference.

The Deduction System

1. The axioms and rules of inference for first-order logic
2. Axioms and inference rules to handle second-order universal generalization and instantiation and second-order existential generalization and instantation

3. The usual axioms for first-order equality
4. The Comprehension scheme: for all formulas $\phi(x_1, \ldots, x_n)$, the following is an axiom: $\exists X \forall x_1 \ldots \forall x_n (X x_1 \ldots x_n \Leftrightarrow \phi(x_1, \ldots, x_n))$
5. Hume's Principle: $\forall X \forall Y (\#X = \#Y \Leftrightarrow X \sim Y)$, where $X \sim Y$ is an abberivation for the formula $\exists Z(\forall x(Xx \Rightarrow \exists! y(Yy \wedge Zxy)) \wedge \forall x(Yx \Rightarrow \exists! y(Xy \wedge Zyx)))$ and where, again, $\exists! x$ is an abberivation for the usual formula meaning "there exists a unique x."

Now it should be noted that when Frege talks about *defining* numbers, what he has in mind is a syntactic procedure. We want to be able to give definitions for the symbols '0', 's', and 'N' using only the language $\mathcal{L}$; we want to be able to translate statements about numbers—in particular the Dedekind-Peano axioms, into $\mathcal{L}$. Unfortunately, this becomes a messy affair which requires expanding $\mathcal{L}$ using λ-calculus or some other method of introducing term-forming operators. For example, we want to define $0 := \#[x : x \neq x]$. (This notation is taken from Boolos and Heck 1998.) Suffice to say that if this is done, then each of the Dedekind-Peano axioms, suitably translated, can be proven in the deduction system given above. (For detailed accounts, see Wright, 1983; Boolos and Heck, 1998; and Zalta, 1998.)

So Frege's theorem is the technical result that the Dedekind-Peano axioms, suitably translated into $\mathcal{L}$—a second-order language with only one non-logical term, #—are theorems of second order logic plus Hume's Principle. There has been substantial philosophical debate over whether or not Frege's theorem amounts to a reduction of arithmatic to logic (For an introduction to the debate, see Hale and Wright, 2001, 2005, Rayo, 2005, and Demopoulos et al, 2005). This question I set aside here, but there are a number of others worth discussing. First, why frame the entire endeavour in terms of an operator # which is applied to concepts and read "the number of ... "? The idea, found in §46 of *Grundlagen* (1884), is given as follows:

> It should throw some light on the matter to consider number in the context of a judgement which brings out its basic use. While looking at one and the same external phenomenon, I can say with equal truth both "It is a copse" and "It is five trees", or both "Here are four companies" and "Here are 500 men". Now what changes here from one judgement to another is neither

> any individual object, nor the whole, the agglomeration of them, but rather my terminology. But that is itself only a sign that one concept has been substituted for another. This suggests as the answer to the first of the questions left open in our last paragraph, that the content of a statement of number is an assertion about a concept. This is perhaps clearest with the number 0. If I say "Venus has 0 moons", there simply does not exist any moon or agglomeration of moons for anything to be asserted of; but what happens is that a property is assigned to the *concept* "moon of Venus", namely that of including nothing under it. If I say "the King's carriage is drawn by four horses", then I assign the number four to the concept "horse that draws the King's carriage". (Trans. by J.L. Austin)

Another question which arises is how Frege's Platonism is engendered by the forgoing analysis of the natural numbers. Unfortunately, much of the transparency is lost when we take the "neo-Fregean" approach given above which adds Hume's principle as a "principle of logic" and takes the operator # as primitive, instead of adding as Frege did (the inconsistent) Basic Law V and giving an explicit definition for # (and *then* deriving Hume's Principle from that definition and Basic Law V).[1] Once this definition is given we get Frege's "full blooded" analysis of numbers: mathematical terms like '0', '1', '2', ... really do refer to objects—specifically they refer to equivalence classes of concepts. Numbers are equivalence classes of concepts.

Most important for Frege's debate with Hilbert, Frege didn't just want to give any old definition of 'natural number', 'successor', and 'zero' from which the Dedekind-Peano axioms could be derived from principles of logic. For Frege, there *really* are numbers: there are, for the terms '0', '$s(0)$', '$s(s(0))$', etc, objects denoted by these terms. In giving his definitions Frege was not only finding—as we would say in contemporary model theory—an interpretation of a set of formal sentences (the Dedekind-Peano axioms) which made them true, but instead was setting out objects (numbers) and

[1] Essentially the definition is as follows. Let $\mathcal{C}$ be the set of all concepts $X, X_1, X_2, \ldots$. Then *equinumerousity*, i.e. $\sim$ as defined above, is an equivalence relation on $\mathcal{C}$ and divides $\mathcal{C}$ into equivalence classes. Then for any X, $\#X$ is the equivalence class of $\mathcal{C}$ containing X. Thus we could say that $\mathbb{N} = \{x : Precedes^*(0, x) \text{ or } x = 0\}$, or, equivalently, we could say $\mathbb{N} = \{x : \exists X \in \mathcal{C} \text{ s.t. } x = \#X\}$, so long as we added some further clause that ruled out concepts X with infinite extensions.

a relation between them (that of predecessor) which the formal sentences were about all along. For comparison, think of everyone's favorite philosopher: Aristotle. I can give a definition which determines the reference of the term 'Aristotle'. Frege famously says that what determines the reference of the term 'Aristotle' is something like "Plato's disciple and the teacher of Alexander the Great". (Frege 1892: ft.2) Once the reference is fixed there are basic facts about Aristotle which we can discover, e.g. that he was born in Stagira, invented logic, and taught that there were four types of causes. The situation is the same with 'zero', 'successor', and 'natural number', the only difference being that once the referent of 'Aristotle' is defined we discover facts about him via empirical discovery, while once the basic terms of arithemtic are defined the basic facts follow from principles of logic and Hume's Principle.[2]

3 Hilbert's Structuralism

There are dozens and dozens of published writings on the above topic. Below I've listed the material cited above. It is certainly a good start for those looking to pursue the topic further.

Hilber's project in the first two chapters of *Grundlagen der Geometrie* (1899) is laid out in the introduction as follows:

> The following investigation is a new attempt to choose for geometry a *simple* and *complete* set of *independent* axioms and to deduce from these the most important geometrical theorems in such a manner as to bring out as clearly as possible the significance of the different groups of axioms and the scope of the conclusions to be derived from the individual axioms. (Trans. by Townsend)

I will not reproduce all of Hilbert's axioms—which he doesn't give in a formal language—but the primitive or undefined terms in the axioms are 'point', 'line', 'plane', 'between', and 'congruent'.[3]

[2]Part of Frege's project was to show that, contrary to Kant, facts about arithmetic were "analytic" and not based on intuition. So if one accepts Frege's point of view on the terms of arithmetic and accepts Hume's Principle as a basic "logical principle", then Frege more or less succeeds.

[3]The Dedekind-Peano axioms recall contained the undefined terms 'zero', 'successor', and 'natural number'. 'Zero' is a constant term—it names a certain object. None of

If Hilbert attempted an analysis of the axioms of geometry along the same lines as Frege's analysis of the axioms of arithmetic, then just as Frege defined a set $\mathbb{N} = \{x : Nx\}$, an element $0 \in \mathbb{N}$ and a function $s : \mathbb{N} \to \mathbb{N}$ and then showed that they satisfied the axioms, Hilbert would have defined the sets $\mathbb{P} = \{x : x \text{ is a point}\}$, $\mathbb{L} = \{x : x \text{ is a line}\}$ and $\mathbb{PL} = \{x : x \text{ is a plane}\}$ and defined relations $B \subseteq \mathbb{P} \times \mathbb{P} \times \mathbb{P}$ and $C \subseteq \{\text{line segments}\} \times \{\text{line segments}\}$ for 'between' and 'congruence'. With all of these defined, Hilbert would have then shown that the axioms held true.

The crux of this approach, of course, would have been that Hilbert's definitions gave a proper, or correct, analysis of the concepts "point", "line", "plane", "between", and "congruence". Such an analysis would presuppose some sort of Platonism about points, lines and planes: there exist such objects which the axioms of geometry are about.

But Hilbert does nothing of the sort, of course. What Hilbert does in chapter 2, after setting out the axioms and proving some theorems in chapter 1, is to show the consistency of the axioms and different independence relations between subsets of the axioms. What he does *not* do is attempt any sort of definitions for the primitive terms like 'point'.

To show consistency, Hilbert shows how we can interpret the term 'point' to mean "ordered pair (u, v) of real numbers" and the term 'line' to mean "ratio $(u : v : w)$ of real numbers such that $u \neq 0 \neq v$".[4] He then interprets the relation of betweenness as some different relation between triples of these pairs of real numbers, and continuing on in this fashion he shows that all his axioms of geometry are true under this interpretation.[5]

That this shows that his axioms are consistent (if the axioms governing the real numbers are) is seen as follows: suppose that a contradiction could be derived from the axioms of geometry. Since all the axioms are true when interpreted as being about pairs of real numbers and whatnot, the axioms of the reals imply the geometric axioms 'with the geometric terms replaced by terms about reals'. So then a contradiction could be derived from the

Hilbert's terms are constants. 'Successor' is a function term—it names a function. Again none of Hilbert's terms are functions. 'Natural number', though, is a unary relation term, just like 'point', 'line', and 'plane'. 'Congruent' is a binary relation term, while 'between' is a 3-place relation term.

[4]Actually, as Hilbert shows one doesn't need to let u, v, w range over all reals, just some subset of them that he specifies.

[5]One must be careful: Hilbert didn't intend these interpretations as providing any sort of definitive analysis of 'point', 'line', etc. They are just one possible interpretation.

axioms of the reals.

To show the independence of certain groups of axioms from others, Hilbert essentially does the same thing. He gives (different) interpretations of the terms 'point', 'line', 'plane', 'between', and 'congruent' on which the axioms in the latter group are true while the axioms in the former group are false.

This shows that the two groups are independent as follows: Say the axioms in the former group could be derived from the axioms in the latter group. Then the axioms reinterpreted under the model which made the latter group true and the former false would also have the same dependence, we could derive the (reinterpted) former group from the (reinterpted) latter group. But this is impossible, since the reinterpted former group is true but the reinterpted latter group is false.

4 The Debate

What did Frege object to about this procedure? There are several ways to lay out the disagreement, but it essentially was a disagreement over the nature of mathematical axioms. One place to start is to describe what Frege would have seen as a "legitimate" proof of the consistency of Hilbert's axioms. Essentially a real consistency proof, Frege thinks, would involve first giving the "proper" or "right" definitions of 'point', 'line', etc, and then showing that the axioms are true of the objects so defined. He says:

> Axioms *do not* contradict each other because they are true; no proof is necessary to establish this fact. Definitions *must not* contradict each other. In definig we must formualate our basic propositions in such a way as to rule out any possibility of contradiction. (Frege 1903)

The deeper idea here is simply that Frege rejected the idea that the terms of the axioms, e.g. 'point', 'line', etc, could be (re)interpreted. For Frege, these terms, like 'Aristotle', have both definite meaning and reference. These references—the points, the number zero, etc—are the purview of mathematics. To reinterpret the axioms is to forget about their mathematical content.

Another way to approach the dispute is from Hilbert's perspective. Why, exactly, did Hilbert see his method as legitemate? The answer is that, at least at the time of his writing his *Grundlagen* (1899), Hilbert rejected Frege's view of mathematical axioms and, certainly, anything like mathematical Platonism. Hilbert specifically rejects Frege's thoughts concerning the need for definitions of terms like 'point' and 'natural number' which provide a "proper" analysis of the terms. This is stated in both his *Grundlagen* and correspondence with Frege:

> The axioms of this group define the idea expressed by the word between, and make possible, upon the basis of this idea, an order of sequence of the points upon a straight line, in a plane, and in space. (Hilbert 1899: §3)

> The axioms of this group define the idea of congruence or displacement. (Hilbert 1899: §6)

> [I]t 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.[6]

Of course, besides these more philosophical concerns there are technical reasons for prefering Hilbert's method which neither he nor Frege knew of at the time. As we know now, there are deep connections between what can be derived from a set of axioms in a formal system and the set of interpretations of those axioms, e.g. in first-order logic a contradiction can be derived from a set of axioms if and only if on every interpretation at least one of the axioms is false. Thus even if you rejected Hilbert's philosophical point of view and held Frege's, Hilbert's methods would afford you a powerful tool for studying the consistency of mathematical axioms.

To summarize, the dispute is essentially over the nature of mathematical axioms and, more broadly, what we're doing when we do mathematics. For

[6] Letter to Frege of December 29, 1899, as excerpted by Frege (ellipsis Hilbert's or Frege's) in Frege, *Philosophical and Mathematical Correspondence*, 1980. Quotation taken from from Blanchette, 2007.

Frege, when we do mathematics we're studying some realm of abstract, mathematical objects—the natural numbers, the points, etc. We need definitions to fix these objects as the referents of the terms in the axioms. As Frege says:

> For a long time an axiom has always been taken to be a thought whose truth is known without being susceptible of proof by a logical chain of reasoning. …
>
> Definition in mathematics usually means a determination of the reference of a word or symbol. Definitions are distinct from all other mathematical propositions in containing a word or symbol which up to then has had no reference; the definition now supplies one. All other mathematical propositions (theorems and ones expressing axioms) must not contain proper names, concept words, relation words or functional symbols, whose reference is not already determined. …
>
> On the other hand, one can never expect basic propositions and theorems to determine the reference of a word or symbol. The rigor of mathematical investigations makes it absolutely imperative that we should not obscure the difference between definitions and all other propositions. (Frege 1903: 3–5)

But what Frege denies in at the end of that quote is precisely what Hilbert affirms: we do not first determine what the referents of the terms are in the axioms and then say that mathematics is about those things, instead we say that mathematics is about whatever satisfies the axioms.

I suggest that this is why Hilbert, at least at this early stage, is clearly a structuralist. The axioms of a mathematical domain provide a "scaffolding or schema" or, even, a *pattern*. One who studies, say, geometry is studying that schema or pattern. What is important in mathematics is not the referents of the terms—to think that is to get caught up in a dubious mathematical Platonism—but instead is the schema or pattern given by the axioms.

References

Blanchette, Patricia. 2007. "The Frege-Hilbert Controversy". In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Revised 2009 edition.

Boolos, George and Heck, Richard. 1998. "Die Grundlagen Der Arithmetik §§82–3". In Matthias Schirn (ed.), *The Philosophy of Mathematics Today*. Oxford University Press, 407–428.

Demopoulos, William and Clark, Peter. 2005. "The Logicism of Frege, Dedekind, and Russell". In Stewart Shapiro (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford University Press, 129–165.

Frege, Gottlob. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Trans. "Concept Script, a formal language of pure thought modelled upon that of arithmetic".

——. 1884. *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Trans. "The Foundations of Arithmetic: A logico-mathematical enquiry into the concept of number".

——. 1892. "Über Sinn und Bedeutung". *Zeitschrift für Philosophie und philosophische Kritik* 100:25–50. Trans. "On Sense and Referene".

——. 1893/1903. *Grundgesetze der Arithmetik*. Jena: Verlag Hermann Pohle. Trans. "The Basic Laws of Arithmetic".

——. 1903. "Über die Grundlagen der Geometrie". *Jahresbericht der Deutschen Mathematiker-Vereinigung* 12:19–24,368–75.

Hale, Bob and Wright, Crispin. 2001. *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Clarendon Press.

——. 2005. "Logicism in the Twenty-first Century". In Stewart Shapiro (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford University Press, 166–202.

Hilbert, David. 1899. *Grundlagen der Geometrie*. Leipzig: Teubner. Trans. "The Foundations of Geometry".

Rayo, Agustín. 2005. "Logicism Reconsidered". In Stewart Shapiro (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford University Press, 203–235.

Wright, Crispin. 1983. *Frege's Conception of Numbers as Objects*. Aberdeen University Press.

Zalta, Edward. 1998. "Frege's Logic, Theorem and Foundations for Arithmetic". In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Revised 2010 edition.

Appendix A: The Language $\mathcal{L}$

Terms:

1. All object variables $x, y, z, \dots, x_1, y_1, z_1, \dots$ are terms.
2. If f is an n-ary function variable and $t_1, \dots, t_n$ are terms, $f(t_1, \dots, t_n)$ are terms.
3. If X is a unary relation variable, $\#X$ is a term.

Formulas:

1. If X is an n-ary relation variable and $t_1, \dots, t_n$ are terms, $Xt_1 \dots t_n$ is formula.
2. If t_1 and t_2 are terms, $t_1 = t_2$ is a formula.
3. If ϕ and ψ are formulas, $\phi \Rightarrow \psi$, $\phi \vee \psi$, $\phi \wedge \psi$, $\phi \Leftrightarrow \psi$ and $\neg\phi$ are formulas.
4. If ϕ is a formula and x an object variable, $\forall x\phi$ and $\exists x\phi$ are formula.
5. If ϕ is a formula X a relation variable, $\forall X\phi$ and $\exists X\phi$ are formula.
6. If ϕ is a formula f is a function variable, $\forall f\phi$ and $\exists f\phi$ are formula.

Appendix B: Models and Frege's Theorem

A *structure* $M = (D, \#^M)$ for $\mathcal{L}$, following the standard semantics for second-order logic, will consist of a set D—the *domain*—and a function $\#^M : \mathcal{P}(D) \to D$ from the powerset of D to D which interprets the function symbol #. An *assignment* α is a function from the object variables of $\mathcal{L}$ to elements of D, from the (n-ary) relation variables of $\mathcal{L}$ to subsets of D^n and from the (n-ary) function variables of $\mathcal{L}$ to functions $D^n \to D$. *Satisfaction* is defined as follows:

1. $M, \alpha \models Xt_1, \dots, t_n$ if and only if $(\alpha(t_1), \dots, \alpha(t_n)) \in \alpha(X)$
2. $M, \alpha \models t_1 = t_2$ if and only if $\alpha(t_1) = \alpha(t_2)$
3. $M, \alpha \models \phi \wedge \psi$ if and only if $M, \alpha \models \phi$ and $M, \alpha \models \psi$; and so on for the other propositional connectives
4. $M, \alpha \models \forall x\phi$ if and only if $M, \alpha' \models \phi$ for all assignments α' which differ from α only on x; and similarly for universal quantification over relation and function variables
5. $M, \alpha \models \exists x\phi$ if and only if $M, \alpha' \models \phi$ for some assignment α' which differs from α only on x; and similarly for existential quantification over relation and function variables

If ϕ is a sentence of $\mathcal{L}$ and $M, \alpha \models \phi$ for all assignments α of M, we call M a *model* of ϕ.

Now we expand $\mathcal{L}$ by adding to its signature the constant 0, the function symbol s and the unary relation symbol N. Call this new language $\mathcal{L}^+$. A structure $M = (D, \#^M, 0^M, s^M, N^M)$ of $\mathcal{L}^M$ will be just like a structure of $\mathcal{L}$, but will also include a designated element $0^M \in D$ denoted by 0, some function $s^M : D \to D$ denoted by s and subset $N^M \subseteq D$ denoted by N. We turn every structure M of $\mathcal{L}$ into a structure of $\mathcal{L}^+$ as follows.

We begin with three preliminary definitions. For each structure $M = (D, \#^M)$ of $\mathcal{L}$, define some subset $Prec^M \subseteq D^2$ so that $(x, y) \in Prec^M$ if and only if there's some subset $A \subseteq D$ and element $z \in A$ such that $y = \#^M(A)$ and $x = \#^M(\{w : w \in X \text{ and } w \neq z\})$. (Frege calls this the *predecessor* relation.) Next, for any subset $A \subseteq D^2$ we define when a set $F \subseteq D$ is *hereditary* in the "A-series": F is hereditary in the A-series if and only if for all $(x, y) \in A$, if $x \in F$ then $y \in F$. Finally, for all $A \subseteq D^2$ we define the set $A^* \subseteq D^2$ by saying that $(x, y) \in A^*$ if and only if for all $F \subseteq D$, (i) for all $z \in D$, if $(x, z) \in A$ then $z \in F$; and (ii) if F is hereditary in the A-series, then $y \in F$. (A^* has come to be known as the *ancestral* of A.)

Now on to the important work: we turn any structure $M = (D, \#^M)$ of $\mathcal{L}$ into a structure of $\mathcal{L}^+$ by setting $0^M = \#(\{x : x \neq x\})$, $s^M(x) = y$ if and only if $(x, y) \in Prec^M$ and setting $N^M = \{x : (0^M, x) \in Prec^*, \text{ or } x = 0^M\}$. We can denote this new structure by M^+.

Frege's Theorem: (Model-theoretic Style) *If $M = (D, \#^M)$ is a structure for $\mathcal{L}$ which satisfies Hume's Principle, then M^+ is a model for the Dedekind-Peano axioms.*