Notre Dame Journal of Formal Logic
Volume 28, Number 1, January 1987

On the Consistency of the First-Order Portion of Frege's Logical System

TERENCE PARSONS*

It is well known that Frege's logical system of his *Grundgesetze der Arithmetik* [1] is inconsistent. However, Peter Schroeder-Heister (in [3]) has speculated that the first-order portion of this system is consistent. On the surface, this is a somewhat surprising conjecture, because Frege's so-called "abstraction" principle is included in the first-order part of his system, and the somewhat similar abstraction principle of (first-order) naive set theory leads quickly to inconsistency. But Frege's abstraction principle is a prima facie weaker principle. Instead of assuming that any formula determines a set which satisfies that formula, it holds only that coextensive formulas must determine the same "courses of values". That is, instead of this:

Set Theorem $(\exists x)(y)(y \in x \equiv A)$, *for any A not containing x*,

we assume (roughly) this:

Frege $(x)(A \equiv B) \equiv \grave{x}A = \grave{x}B$, *for any A, B.*[1]

It is well known that if quantification over functions is admitted into Frege's system (as Frege himself did) then it is possible to define an analogue of set membership, and the abstraction principle of naive set theory can be shown to follow from Frege's abstraction principle.[2] Russell's paradox quickly follows. But it is not obvious how to do this within the first-order portion of Frege's system. The first goal of this paper is to show that this cannot be done. Schroeder-Heister's conjecture is correct: the first-order portion of Frege's system is consistent.

The second goal of this paper is to explore the significance of the model-construction technique sketched herein for Frege's claims about the arbitrariness of the identification of truth-values with courses of values. Although Schroeder-Heister has shown that Frege's claims on this topic are false, there are some closely related claims that are true and interesting.

*Work on this paper was stimulated by [3], and many of the ideas and techniques used here originated there.

Received November 2, 1984

1 Consistency of the system To understand the following details, we need to keep in mind Frege's beliefs that: (1) truth-values are objects, not to be distinguished ontologically from other objects, and (2) terms which denote truth-values can occur syntactically in the same places that other names of objects can. Since sentences denote truth-values, this means that sentences can occur in all of the places where we would normally expect names of objects to occur, such as flanking the identity sign. This allows Frege, e.g., to use the identity sign for the material biconditional.

1.1 Syntax of the formal language L In what follows I will use "$[X^Y_Z]$" as an abbreviation for "the result of replacing each occurrence of Y in X by an occurrence of Z". I will generally use signs of the object-language as metalinguistic names of themselves. I follow Frege's custom of using Gothic letters ($\mathfrak{a}_1, \mathfrak{a}_2, \ldots$) for variables bound by the universal quantifier, and Greek letters ($\epsilon_1, \epsilon_2, \ldots$) for variables bound by the course-of-values abstraction symbol "’".

The syntax of the first-order part of Frege's system is this:

The object parameters of L are: $x_1, x_2, x_3, \ldots$

Every object parameter of L is a complete name of L.

The function parameters of L are: $f_1, f_2, f_3, \ldots$

If f_n is a function parameter of L, and A is a complete name of L, then this is a complete name of L: $f_n(A)$.

If A and B are complete names of L, so are the following:

Using the horizontal:	$-A$
Using negation:	$\neg A$
Using the conditional:	$(A \rightarrow B)$
Using identity:	$(A = B)$
Using universal quantification:[3]	$(\mathfrak{a}_i)[A^{x_i}_{\mathfrak{a}_i}]$
Using course-of-values abstraction:	$\acute{\epsilon}_i[A^{x_i}_{\epsilon_i}]$

1.2 Axiomatics I assume that we are given a complete set of rules and axioms for the first-order predicate calculus with identity, expressed within Frege's system. I will call this the "logical" system. In addition, I will be discussing Frege's so-called "abstraction" principle governing course-of-values names. This principle states that for any complete names A and B of L the following is to be an axiom:

$$(\mathfrak{a}_i)([A^{x_n}_{\mathfrak{a}_i}] = [B^{x_n}_{\mathfrak{a}_i}]) = (\acute{\epsilon}_j[A^{x_n}_{\epsilon_j}] = \acute{\epsilon}_k[B^{x_n}_{\epsilon_k}]).[4]$$

If we think of the formulas A and B as expressing functions, this principle tells us that A and B express coextensive functions if and only if the courses-of-values of those functions are the same. Our problem is to see whether there is a model of Frege's logical system in which this abstraction principle is true.

1.3 Interpretations Suppose that U is a set (a "universe"), and that t and f are distinct members of U. Intuitively, U represents the class of all Fregean objects, and t and f are those objects that Frege calls "The True" and "The False". I will call σ a *basic assignment* over U if σ is any assignment of members of U to the object parameters of L, and of members of U^U to the function parameters of L. Then an *interpretation I over U of* (a portion of) L is to be a function which, given any basic assignment σ over U, produces an assignment I^σ of members of U of *all* of the complete names of (that portion of) L, and an assignment of members of U^U to all of the function parameters of (that portion of) L, where it is understood that I^σ agrees with σ on all of the parameters of L.

If I is an interpretation and σ a basic assignment, we say that I^σ makes A true if $I^\sigma[A] = t$, and makes A false if $I^\sigma[A] = f$. When we say that I alone makes A true, we mean that I^σ makes A true for any basic assignment σ. There is no presumption so far that an interpretation and a basic assignment make even the logical theorems of L true. Our task will be to show, first, how to produce an interpretation of that part of L which does not contain any course-of-values names, and which makes true all of the logical theorems of that part of L. Then we will show how to extend any such interpretation to the rest of L, including course-of-values names, so as to also make true the abstraction principle.

Suppose that I is any interpretation over U of some portion L' of L, which may contain all, some, or none of the course-of-values names of L. Assume also that L' is *syntactically grounded*, in the sense that for every name A occurring in L', if B occurs in the syntactic rule that generates A, then B is also in L'. Then corresponding to I is a unique interpretation I^* of L', determined by the following conditions (which hold for any basic assignment σ):

$I^{*\sigma}[x_n] = I^\sigma[x_n]$ for every object parameter x_n.

$I^{*\sigma}[f_n] = I^\sigma[f_n]$ for every function parameter f_n.

$I^{*\sigma}[\grave{\epsilon}_i A] = I^\sigma[\grave{\epsilon}_i A]$ for every course-of-values name of L for which I^σ is defined.

$$I^{*\sigma}[f_n(A)] = I^{*\sigma}[f_n](I^{\sigma^*}[A])$$

$$I^{*\sigma}[-A] = \begin{cases} t \text{ if } I^{*\sigma}[A] = t \\ f \text{ otherwise} \end{cases}$$

$$I^{*\sigma}[\mathrel{-\!\!\!\!\top}\! A] = \begin{cases} t \text{ if } I^{*\sigma}[A] \neq t \\ f \text{ otherwise} \end{cases}$$

$$I^{*\sigma}[(A \to B)] = \begin{cases} t \text{ if } I^{*\sigma}[A] \neq t \text{ or } I^{*\sigma}[B] = t \\ f \text{ otherwise} \end{cases}$$

$$I^{*\sigma}[(A = B)] = \begin{cases} t \text{ if } I^{*\sigma}[A] = I^{*\sigma}[B] \\ f \text{ otherwise} \end{cases}$$

$$I^{*\sigma}[(\mathfrak{a}_k)A] = \begin{cases} t \text{ if } I^{*\sigma[x_k/u]}[A^{\mathfrak{a}_k}_{x_k}] = t \text{ for every } u \in U, \\ f \text{ otherwise} \end{cases}$$

(Note: $\sigma[x_k/u]$ is that assignment which is exactly like σ except that it assigns u to x_k. Recall that by the syntactic rules given above, x_k will never occur in $(\mathfrak{a}_k)A$.)

1.4 The model Suppose that we have an interpretation I over some infinite set U for that portion of L which contains no course-of-values names.[5] I assume that it is clear that I^* makes true all of the logical theorems of the system. This part of Frege's system is so much like the ordinary (first-order) predicate calculus with identity that conventional modern methods apply. Given such an interpretation, the task is to show how to extend I to the whole of L in such a manner that the abstraction principle is satisfied.

First, define the rank of any course-of-values name $\grave{\epsilon}_i A$ to be 1 if A contains no course-of-values names, and to be $1 + rank[B]$ if B is a course-of-values name in A whose rank is at least as high as that of any other such name in A.

Now order the course-of-values names of L by rank, and, within rank, in some arbitrary manner (of order-type ω). I will use $\grave{\epsilon}_i A_{n,m}$ to denote the mth course-of-values name of rank n.

Next, choose any countably infinite subset of U (perhaps containing all of U), and order this subset in any way you like into a countable sequence of countable sequences. I will use $u_{n,k}$ to denote the kth member of the nth sequence. It is understood that if $n \neq n'$ or $k \neq k'$ then $u_{n,k} \neq u_{n',k'}$.

We will show how to assign to each name $\grave{\epsilon}_i A_{n,m}$ some $u_{n,k}$ as its referent (relative to each basic assignment σ). This will be done in stages. Given our initial interpretation I, we define interpretations $I_{n,m}$ in a step-by-step fashion, as follows:

Basis $I^{\sigma}_{1,0} = I^{*\sigma}$ (for each σ)

Successor Step We first extend $I_{n,m}$ to $I'_{n,m}$ by stipulating the value of $I'^{\sigma}_{n,m}[\grave{\epsilon}_i A_{n,m+1}]$, for any σ, as follows:

> If there is some name B for which $I^{\sigma}_{n,m}[\grave{\epsilon}_k B]$ is already defined, and for which $I^{\sigma[x_j/u]}_{n,m}[A^{\epsilon_i}_{x_j}] = I^{\sigma[x_j/u]}_{n,m}[B^{\epsilon_k}_{x_j}]$ for every $u \in U$, then $I'^{\sigma}_{n,m}[\grave{\epsilon}_i A_{n,m+1}] = I^{\sigma}_{n,m}[\grave{\epsilon}_k B]$. Otherwise, $I'^{\sigma}_{n,m}[\grave{\epsilon}_i A_{n,m+1}] =$ the next unused $u_{n,k}$ of rank n.

(We call $u_{n,k}$ "unused" if it has not yet been assigned to any course-of-values name. Also, we assume for this account that the subscript j of x_j is chosen so as not to occur as a subscript of a Greek or Gothic letter in either A or B.)

Finally, we set $I^{\sigma}_{n,m+1} = I'^{*\sigma}_{n,m}$.

Limit Step We set $I^{\sigma}_{n+1,0} = \bigcup_{k} I^{\sigma}_{n,k}$.

(It may be verified that $I^{\sigma}_{n,m}$ is always a subfunction of $I^{\sigma}_{n,m+1}$, so the limit step does yield a single-valued function.)

To get the desired interpretation, we now set $I^{\sigma}_{\omega} = \bigcup_{n,m} I^{\sigma}_{n,m}$. Then I_{ω}, as so defined, is the desired interpretation.

Theorem *All instances of the abstraction schema (as well as all of the other logical theorems) are true under* I_{ω}.

This theorem may be proved by a straightforward induction on the ordering used in the construction.[6]

2 The arbitrariness of the identification of truth-values with courses-of-values In the first few sections of the *Grundgesetze* Frege achieves a certain elegance by "identifying" the two truth-values t and f with the courses of values named by $\grave{\epsilon}(-\epsilon)$ and $\grave{\epsilon}(\epsilon = \mathrel{-\!\top\!-}(\mathfrak{a})(\mathfrak{a} = \mathfrak{a}))$. In defense of this policy he adopts a conventionalist stance, and defends his choice by claiming that such identification is arbitrary. Schroeder-Heister has shown that this is an overgeneralization (see (3) below), and he asks about the limits of such identifications. The construction of Section 1 sheds some light on this issue.

Suppose that we have selected which objects (which members of U) are to be the truth-values, and that we have fixed on an interpretation and a basic assignment which jointly establish denotations for all of the names of L that do not contain course-of-values names. How does this constrain which objects must be selected to be which courses of values? A Fregean moral of the construction given above is that it hardly constrains it at all. In particular, we have the following results:

(1) It is always possible to find a model for the entire language in which neither of the truth-values is the referent of any course-of-values name relative to any σ. Just use the construction given above, leaving both t and f out of the set of objects chosen to be courses of values.

(2) It is always possible to make Frege's choice, and to identify t as the referent of $\grave{\epsilon}(-\epsilon)$ and f as the referent of $\grave{\epsilon}(\epsilon = \mathrel{-\!\top\!-}(\mathfrak{a})(\mathfrak{a} = \mathfrak{a}))$. Just choose these names as the first and second names of rank 1, and pick $u_{1,1}$ to be t and $u_{1,2}$ to be f. (The only thing that needs verifying here is that $I^{\sigma[x/u]}[-x]$ will disagree with $I^{\sigma[x/u]}[x = \mathrel{-\!\top\!-}(\mathfrak{a})(\mathfrak{a} = \mathfrak{a})]$, for some $u \in U$. In fact, they will always disagree for $u = t$.) Reversing the choice is also always possible.

(3) It is not possible, in general, to select any arbitrary pair of course-of-values names and assume that the first may be assigned t as its denotation and the second assigned f. This is part of what Schroeder-Heister has shown. For example, it is never possible to arrange things so that $\grave{\epsilon}(\epsilon = \mathrel{-\!\top\!-}(\mathfrak{a})(\mathfrak{a} = \mathfrak{a}))$ denotes t *and* $\grave{\epsilon}_1(\epsilon_1 = \mathrel{-\!\top\!-}\grave{\epsilon}(\epsilon = \mathrel{-\!\top\!-}(\mathfrak{a})(\mathfrak{a} = \mathfrak{a})))$ denotes f. If, for example, the construction given above is arranged so that the former course-of-values name denotes t, then the latter will be forced to denote t as well. (Note that *in general* the two names may receive different denotations.)

(4) However, there is an ontological analogue of this principle that is true, and this is the point that I think Frege should have made. Let me say that a course-of-values name $\grave{\epsilon}A$ *signifies the function* g (relative to I and σ) if and only if $I^{\sigma[x/u]}[A^{\epsilon}_{x}] = g(u)$ for every $u \in U$. Then, given *any* two functions g and h which are members of U^U, it is possible to select t as the course-of-values of g and f as the course-of-values of h. More precisely, it is possible to find an assignment σ and two course-of-value names such that the first name signifies g and denotes t (relative to I and σ), and the second signifies h and denotes f (relative to I and σ). (Just pick a σ which assigns g to f_1 and h to f_2, and choose the course-of-values names to be $\grave{\epsilon}f_1(\epsilon)$ and $\grave{\epsilon}f_2(\epsilon)$, and then carry out the construction as in (2) above.) The arbitrariness that so impressed Frege has its source in the arbitrariness of the connection between functions and their courses of values in his system. His abstraction principle requires there to be a 1-1 correlation between functions and objects, but it puts absolutely no further constraints on this correlation.[7] This is a theme of "Über Begriff und Gegenstand" [2], in which he holds that each concept is represented by some object, but in which he never says *which* objects represent which concepts.

(5) The point just made depends on a special feature of our language L: the presence in L of function parameters. These are needed in order to make sure that the chosen functions g and h can be signified by some name in the language. For if they could not be so signified, then the construction given in Section 1 would not apply to them. It is worth asking about a language that may have been closer to that which Frege had in mind, namely, that portion of L which contains no parameters at all. The only symbols present are the logical ones, along with bound Gothic and Greek letters. This language is limited in expressive power in a certain way. Crudely put, its formulas with one free variable each can only distinguish among the truth values and the objects named by course-of-values names; the rest are all treated alike. As a result, the complete names not containing course-of-values names have their references logically fixed, and it is totally determined which functions will be signified by course-of-value names of rank 1. In particular, if two such course-of-value names signify the same function relative to one interpretation they signify the same function relative to any interpretation, and if they signify different functions in one interpretation then they do so in all. So in such a restricted language, any two nonsynonymous course-of-values names *of rank 1* may be selected as designating the two truth-values (in either order).[8] Notice that Frege's own choice is an example of this generalization. This result does not generalize to any higher ranks, for Schroeder-Heister already shows how to get a counterexample using names of ranks 1 and 2, and any course-of-values name may have its rank artificially boosted by incorporating in it superfluous course-of-values names.

NOTES

1. In this formula $\grave{x}A$ and $\grave{x}B$ are taken to denote the "courses of values" related to formulas A and B. Frege's principle is actually broader than the one given, since he has courses of values corresponding to any open name, not just to the names that correspond to formulas in the modern sense.

2. We can define membership as follows:

$$x \in y \equiv (\exists f)[f(x) \;\&\; y = \grave{z}f(z)].$$

You get Russell's paradox by asking whether $\grave{z}(\sim z \in z)$ is a member of itself, using the definition and Frege's abstraction principle.

3. As given, the rule for the universal quantifier prohibits the formation of a complete name in which some object parameter has the same subscript as a Gothic letter in whose scope it falls, and the rule for course-of-values names prohibits the formation of such a name in which an object parameter has the same subscript as that of some Greek letter in whose scope it falls. These are not important restrictions. Note that it might be truer to Frege's views about the "gappiness" of incomplete names to disallow vacuous quantification; this is not, however, relevant to the points under discussion.

4. In order to prevent "capturing" we need to restrict this schema to instances in which x_n does not already occur within the scope of ϵ_j in A, or ϵ_k in B, or $\mathfrak{a}_i$ in either A or B.

5. If U is not infinite then no model will be possible, since Frege's abstraction principle, together with his first-order logical principles, forces an infinite domain. (That is, it does this if we assume that $t \neq f$. Otherwise it is easy to provide a model for his system with a 1-element domain, even if the system is inconsistent.)

6. In discussing Frege's system I have left out his definite description operator. This made the construction easier to follow. If the operator is included, the following additions should be made:

   First, for any course of values name $\acute{\epsilon}A$, we add to the symbolism the definite description $(\iota\epsilon)A$.

   Next, in giving the construction, we assume initially that L lacks definite descriptions as well as course of values names.

   Then, every time we stipulate the value of $I'^{\sigma}_{n,m}[\epsilon_\iota A_{n,m+1}]$, we follow this with the stipulation that $I'^{\sigma}_{n,m}[\iota\epsilon_\iota A_{n,m+1}]$ is to be that unique member u of U which is such that $I^{\sigma[x_k/u]}_{n,m}[A^{\epsilon_\iota}_{x_k}] = t$, if there is such a u, and otherwise it is to have the same denotation as $I'^{\sigma}_{n,m}[\acute{\epsilon}_\iota A_{n,m+1}]$. (This is just a formal statement of Frege's own stipulation.)

7. In the first-order fragment there must be a 1-1 correlation between named objects and signified functions. In the higher-order version the abstraction principle may be quantified, and it then requires a 1-1 correlation between all objects and *all* functions, in violation of Cantor's theorem.

8. Let $\acute{\epsilon}_j[A]$ and $\acute{\epsilon}_k[B]$ be two course-of-values names. Let x_n be any object parameter which does not occur in either name. Then I call the names *synonymous* if for every I and σ:

$$I^{*\sigma[x_n/u]}[A^{\epsilon_j}_{x_n}] = I^{*\sigma[x_n/u]}[B^{\epsilon_k}_{x_n}] \text{ for every } u \in U.$$

The advertised result may be proved as follows. First, we have two lemmas:

(1) Suppose that A is any complete name that does not contain any course-of-values names. Then it is easy to show inductively that for any interpretations I and I', and basic assignment σ, that:

⟨1⟩ $I^{*\sigma}[A] = I'^{*\sigma}[A]$.

(2) It can also be established that if σ and σ' do not differ in what they assign to the parameters of A, then for any interpretation I:

⟨2⟩ $I^{*\sigma}[A] = I^{*\sigma'}[A]$.

Now suppose that $\acute{\epsilon}_j A$ and $\acute{\epsilon}_k B$ contain no parameters and no embedded course-of-values names, and suppose that they are not synonymous. This means that for some particular I, σ, and u:

⟨3⟩ $I^{*\sigma[x_n/u]}[A^{\epsilon j}_{x_n}] \neq I^{*\sigma[x_n/u]}[B^{\epsilon_k}_{x_n}]$.

But, since A and B contain no parameters, principle ⟨2⟩ lets us generalize this to any interpretation I. And since they contain no course-of-values names, principle ⟨1⟩ lets us generalize this to any σ. That is, we have that:

⟨4⟩ For any I and any σ there is some $u \in U$ (actually the same u in each case, though this is not needed) such that:

$$I^{*\sigma[x_n/u]}[A^{\epsilon_j}_{x_n}] \neq I^{*\sigma[x_n/u]}[B^{\epsilon_k}_{x_n}].$$

But this is exactly the condition that is needed to allow us to use $\acute{\epsilon}_j[A]$ and $\acute{\epsilon}_k[B]$ in the construction in part 1 so as to assign them the two truth-values, as in (2) above.

REFERENCES

[1] Frege, Gottlob, *Grungesetze der Arithmetik*, Jena, 1893-1903. Relevant portion of Vol. I. Translated by M. Furth as *The Basic Laws of Arithmetic*, University of California Press, Berkeley, 1965.

[2] Frege, Gottlob, *Über Begriff und Gegenstand*, *Vierteljahrsschrift für Wissenschaftliche Philosophie*, Vol. 16 (1892), pp. 192–205. Translated by P. Geach as "On concept and object," in P. Geach and M. Black (eds.) *Translations from Philosophical Writing's of Gottlob Frege*, Blackwell, Oxford, 1952.

[3] Schroeder-Heister, Peter, "A model-theoretic reconstruction of Frege's permutation argument," *Notre Dame Journal of Formal Logic*, vol. 28, no. 1 (1987), pp. 69–79.

Department of Philosophy
University of California, Irvine
Irvine, California 92717