

Notes for 9/6: the inconsistency in Frege's system (and some hint of what the system is)

Randall Holmes

September 9, 2020

In this lecture, I plan to present the proof of a contradiction in Frege's system.

A prerequisite for this is that I actually give some impression of his axioms and rules of inference. I would also like to give some hint of his mathematical aims.

For logic, I reserve the right to use familiar rules of first-order logic for Frege's propositional connectives and quantifiers, because in fact they are valid for his system (which is the first presentation of this logic in fully modern terms). This much of Frege's system is very much alive.

But some new features do need to be presented.

quantifiers over functions: One is that Frege allows quantification over functions. He in fact makes essential use of quantification over functions in defining the application of a course of values (which is an object) to another object. The logic of his quantifiers over functions is standard.

his axiom (definition?) of equality: For completeness I'll state his axiom for equality:

$$G(x = y) \rightarrow G((\forall F : F(y) \rightarrow F(x)))$$

What he is saying here is that $x = y$ can be replaced in any context with $(\forall F : F(y) \rightarrow F(x))$. You might wonder why he doesn't just define $x = y$ as $(\forall F : F(y) \rightarrow F(x))$...well in fact, he is doing that, but his official procedure for stating a definition requires a prior understanding

of equality. So he cannot present this as a definition, and does it as an axiom.

You also might wonder why it isn't a biconditional, and also (I wonder this) why it isn't $(\forall F : F(y) \rightarrow F(x))$.

Notice that $x = y$ is equivalent to $(\forall F : F(y) \rightarrow F(x))$, which implies $F(y) \rightarrow F(x)$ for any particular F . For example, replace F with $\neg H(\mathbf{X})$. Thus we get $\neg H(y) \rightarrow \neg H(x)$ so we actually get $H(x) \rightarrow H(y)$. H was arbitrary so we have shown $(\forall H : H(x) \rightarrow H(y))$ (or, renaming bound variables) $(\forall F : F(x) \rightarrow F(y))$.

In other words, we have shown that $(\forall F : F(y) \rightarrow F(x))$ implies $(\forall F : F(x) \rightarrow F(y))$, too, and we write $(\forall F : F(x) \leftrightarrow F(y))$

functions and their courses of values, axiom 5: Functions for Frege are not objects (they are as it were of a distinct logical type). But with each function (for example, $f = \mathbf{X}^2 + \mathbf{X} + 1$) he associates a course of values, which he writes something like $\epsilon(\epsilon^2 + \epsilon + 1)$ and which we might write $(x \mapsto x^2 + x + 1)$. These courses of values are objects that he postulates.

His axiom 5 says this, in modern notation:

$$(\forall FG : (\forall x : F(x) = G(x)) \leftrightarrow (x \mapsto F(x)) = (x \mapsto G(x)))$$

Here F and G are function variables, and notice that they are universally quantified over.

This tells us that a pair of functions have the same extension iff the corresponding courses of values are equal. He does not say that the functions are equal: for him, equality does not apply to functions, but only to objects.

applying courses of values defined: We need (for practical reasons if we are going to do anything in this system) to define “function application” in the case where the “function” applied is actually a course of values. Frege’s notation for this isn’t easily reproduced: I will define notation $f'x$, following Russell, for the result of applying the function whose course of values f is to x .

$y = f'a$ is defined as

$$(\exists F : f = (x \mapsto F(x)) \wedge y = F(a))$$

Frege has a notation \backslash which is in effect a description operator: $\backslash \epsilon(F(\epsilon))$ is defined (for a concept F) as the unique object y such that $F(y)$, if there is such a unique object. In our notation, the unique y such that $F(y)$ can be written $\backslash(x \mapsto F(x))$. Using this notation, we can define $f'x$ as

$$\backslash(y \mapsto (\exists F : f = (x \mapsto F(x)) \wedge y = F(x)))$$

discussion of the description operator: Notice the essential role played here by existential quantification over functions.

The defining axiom of \backslash , by the way, is $(\forall a : \backslash(x \mapsto x = a) = a)$

This works as intended because if F is a concept and there is a unique y such that $F(y)$, it follows that $(\forall x : F(x) = (x = y))$. [we might more naturally write this as $(\forall x : F(x) \leftrightarrow (x = y))$, but \leftrightarrow coincides with $=$ on truth values]. It then follows by Axiom 5 that $(x \mapsto F(x)) = (x \mapsto x = y)$. Then by substitution of equals for equals it follows that $\backslash(x \mapsto F(x)) = \backslash(x \mapsto x = y) = y$. If \backslash is applied to something other than a course of values of a concept which holds of one and only one object, we don't care what it represents, but Frege does give an intended default meaning (though he does not enforce it in his axioms).

verifying the definition of application of courses of values: It is worth observing that the conditions for application of \backslash will hold in the definition of $f'x$ if f actually is a course of values.

$(x \mapsto G(x))'a = \backslash(y \mapsto (\exists F : (x \mapsto G(x)) = (x \mapsto F(x)) \wedge y = F(a)))$ and in fact $(\exists F : (x \mapsto G(x)) = (x \mapsto F(x)) \wedge y = F(a))$ is a predicate true of $G(a)$ and only true of $G(a)$:

it is true of $G(a)$ because there is a function F (G itself) such that $(x \mapsto G(x)) = (x \mapsto F(x)) \wedge G(a) = F(a)$, that is, $(x \mapsto G(x)) = (x \mapsto G(x)) \wedge G(a) = G(a)$.

Suppose it is true of y : then we show $y = G(a)$. We are given an F such that $(x \mapsto G(x)) = (x \mapsto F(x)) \wedge y = F(a)$. By axiom 5 and

$(x \mapsto G(x)) = (x \mapsto F(x))$ we get $(\forall x : F(x) = G(x))$, so $F(a) = G(a)$, so $y = F(a) = G(a)$.

So it follows that $(x \mapsto G(x))'a = \setminus (y \mapsto (\exists F : (x \mapsto G(x)) = (x \mapsto F(x)) \wedge y = F(a))) = G(a)$: our course of values application operation works as intended.

Developing arithmetic in Frege's system: All of this is essential for Frege to do the mathematics he wants to do.

Before I reveal how this all breaks down, I want to give some idea of what Frege was trying to do. His mathematical aims were fairly modest in modern terms: he wanted to give an account of arithmetic. I am not going to present all the details, but I will give some basic definitions.

the basic idea: He wanted to in effect define each natural number n as (the course of values of) a concept under which (courses of values of) concepts true of exactly n elements would fall.

zero: So, 0 could be defined as $(x \mapsto x = (y \mapsto y \neq y))$. There is exactly one course of values shared by all concepts under which nothing falls (an application of axiom 5), and to be the course of values of a concept under which 0 objects fall is to be exactly that course of values.

successor started : Suppose that K_n has been defined as the course of values of the concept under which fall exactly the courses of values under which n objects fall.

Notice that this means that $K_n'a$ is equivalent to “there are exactly n objects x such that $a'x$ ”.

Consider $(\exists a : K_n'a \wedge (\exists b : \neg a'b \wedge (\forall u : k'u \leftrightarrow a'u \vee u = b)))$.

This predicate says of an object k that there is a such that $a'x$ is true for exactly n objects x , and a b for which $a'b$ is not true, such that $k'u$ is true for any u iff $a'u$ is true or $u = b$. This is true if and only if there are exactly $n + 1$ objects u for which $k'u$ is true.

Define K_{n+1} as

$$(k \mapsto (\exists a : K_n'a \wedge (\exists b : \neg a'b \wedge (\forall u : k'u \leftrightarrow a'u \vee u = b))))),$$

and define the successor operation σ (add one) as

$$(n \mapsto (k \mapsto (\exists a : n'a \wedge (\exists b : \neg a'b \wedge (\forall u : k'u \leftrightarrow a'u \vee u = b))))).$$

This is all notationally mysterious, but it hinges on the idea that a set k has $n + 1$ elements if and only if there is a set a with n elements and an object b not in a , and $k = b \cup \{a\}$ (the elements of k are the elements of a , and in addition b).

the concept “natural number: Then Frege can define the concept “ \mathbf{X} is a natural number” as

$$(\forall I : I'0 \wedge (\forall xy : I'x \wedge (\sigma'x)'y \rightarrow I'y) \rightarrow I'\mathbf{X}) :$$

an object is a natural number iff it has every inductive property. And this definition gives one the powerful principle of mathematical induction to play with, and one can both prove the other Peano axioms for arithmetic and define addition and multiplication with Frege’s toolkit (I will not give details here).

remarks: A couple of comments on my approach: I don’t promise that my details are always the same as Frege’s. His notation is so different that it is maddening to check that one has things exactly right. I also think that he may start the natural numbers with 1 instead of 0, but the definition of 0 and of natural numbers subsuming 0 is perfectly feasible in Frege’s logic. I didn’t actually need to define successor as a double course of values; I realized I could actually define the successor function, but on reflection I left it this way because I like showing you the double course of values construction.

representing 2-argument functions or relations as double courses of values:

I didn’t lecture this, but it is a nice observation.

Just as he uses courses of values to turn concepts or general functions of one argument into objects, he can use “double courses of values” to turn relations or general functions of two arguments into objects. This is actually quite a sophisticated logical maneuver, usually incorrectly attributed to the much later worker Haskell Curry, and so called “currying”. With a function $R(\mathbf{X}, \mathbf{Y})$ associate $(y \mapsto (x \mapsto R(x, y)))$. Notice that this is a very different treatment of functions of two variables than we are used to. Addition for example is “objectified” as **plus** = $(y \mapsto (x \mapsto x + y))$. Observe that **plus**’2 = $(x \mapsto x + 2)$. Addition is identified with (the course of values of) the function of one argument which sends

each number n to the (course of values of the) function “add n ”. Anachronistically speaking, the advantage of this treatment is that we do not need to supply ordered pairs as arguments. A numerical sum is obtained by two applications: $(\text{plus}'2)'3 = 3 + 2 = 5$.

disaster strikes (the paradox of Russell): But at this point tragedy intervenes (quite literally, for Frege the man).

We can now consider the function $R = \neg\mathbf{X}'\mathbf{X}$. Define r as $(x \rightarrow \neg x'x)$, the course of values of the function R .

By the theorem we proved about course of values application, $r'r = (x \mapsto \neg x'x)'r = (x \mapsto R(x))'r = R(r) = \neg r'r$. That is, if $r'r$ is the True, it is the False, and if it is the False, it is the True. Frege’s logic cannot accommodate this any better than ours (because Frege’s logic, without axiom 5, is basically our first-order logic).

I haven’t been able to find a good text for the correspondence between Frege and Russell, but one thing that I do know is that Russell’s original version of the paradox made no sense, because in fact Frege *does* draw the distinctions of logical type which are one of the defenses against the “paradoxes of set theory”. Russell spoke initially of a predicate of predicates ϕ defined by $\neg\phi(\phi)$. But Frege doesn’t allow $\phi(\phi)$: he recognizes that functions and objects are of essentially different sorts and he doesn’t allow $F(F)$ because this notation tells us that what is outside the parentheses is a function, what is inside is an object, so the same letter cannot refer to both.

Drawing this kind of distinction is the heart of Russell’s approach to averting the paradoxes and Frege already draws it!

His system collapses not because he does not know the difference between functions and objects, but because he *explicitly* postulates the correspondence between all functions (at least, all first-order functions of one argument) and certain objects (the courses of values) expressed in axiom 5. I would say that this is not a logical truth at all, and Frege’s logicism is not refuted by its failure.