Repairing Frege's Logic

M. Randall Holmes

August 5, 2018

This note makes an observation which has been made by others, though perhaps not in the same way. The observation is that the logical system of Frege is readily salvaged by imposing the discipline of stratification originally proposed in Quine's 1937 paper New Foundations. The resulting system is consistent because it is essentially equivalent to the system NFU proposed by Jensen in 1969 and shown to be consistent (see [**?**]). (Nino Cocchiarella for example, has given a full formal description of a development of Frege's system with paradoxes removed by imposing stratification in [**?**]). The details of how stratification is best defined in this context are worth spelling out precisely. It should be noticed that we are making some strategic decisions as we do this which could have been made differently: this is not the only way to formulate a stratified version of Frege's foundational system; the reader can for example compare what we do with what Cocchiarella did. There is then a further question, of a philosophical rather than a mathematical nature: is it possible to give a justification for the repair in philosophical terms more satisfying than the simple observation that it works as a piece of mathematics?

It should be noted that this approach is quite different from the recent neo-Fregean program abandoning Axiom V and adopting Hume's Principle. The systems resulting from the latter approach are far weaker.

The payoff is considerable. The entire development of the Grundgesetze works. One does need to add an axiom of infinity to the purely logical axioms, as NFU does not prove infinity. NFU + Infinity is a fully impredicative system (with the same strength as Principia with the axiom of reducibility, or Zermelo set theory with comprehension restricted to bounded formulas), more than adequate for all of classical mathematics, and certainly adequate for arithmetic.

We present axioms for propositional logic adapted from Frege's Begriffschaft to modern notation.

P1 $P \to (Q \to P)$

P2 $(R \to (Q \to P)) \to ((R \to Q) \to (R \to P))$

P3 $(S \to (Q \to P)) \to (Q \to (S \to P))$

P4 $(Q \to P) \to (\neg Q \to \neg P)$

P5 $\neg\neg P \to P$

P6 $P \to \neg\neg P$

Along with

Rule 1 (modus ponens): From P and $P \to Q$ deduce Q.

this is an adequate set of rules for propositional logic (with suitable definitions of the other usual logical operators). Any system of classical propositional logic will do, of course.

Axioms for identity and quantification follow:

I1 $a = b \to (Fa \to Fb)$

I2 $a = a$

Q1 $(\forall x.Fx) \to Fa$

Further, we have

Rule 2 (universal generalization): if we can prove Fa, where a is an arbitrary constant about which we have assumed nothing, we can deduce $(\forall x.Fx)$.

Here the expression Fx is to be understood as representing any formula in which the variable x appears. In the formalization below, F is taken to represent any second order term.

Now we have to introduce more detail. What we have so far is a fairly standard presentation of first-order logic with identity. Frege's full system is second-order logic with an additional abstraction operator for converting second-order terms to first-order terms. Our approach requires more care with syntax.

First order terms of our language (representing objects) are of the following forms:

variables: First order variables x, with each variable having an associated integer type $\texttt{type}(x)$ Notice here and throughout that `type` operates on syntax, not the objects represented by the syntax. You may suppose the argument enclosed in quotes, but we will not do this.

constants: We further provide constants a, to which no type needs to be assigned,

abstraction terms: We further provide abstraction terms $\hat{x}A(x)$, where $A(x)$ is a formula (which may or may not include the variable x). The type of an abstraction term is governed by $\texttt{type}(\hat{x}A(x)) = \texttt{type}(x)+1$. Occurrences of x in $A(x)$ are bound in $\hat{x}A(x)$.

Formulas of our language are of the following forms:

equations: Equations $t = u$, where t and u are first order terms and either at least one of t and u contains no free variables (of either order) or $\texttt{type}(t) = \texttt{type}(u)$ (this is part of the stratification discipline).

applications: Second order terms applied to first order terms: Fx, where F is a second-order term, x is a first-order term, and either at least one of F and x contains no free variables (of either order) or $\texttt{type}(F) = \texttt{type}(x) + 1$.

implications: Implications $P \rightarrow Q$ where P and Q are formulas.

negations: Negations $\neg P$ where P is a formula.

first order quantifications: Universal sentences $(\forall x.A(x))$ where $A(x)$ is a formula (which may or may not actually contain x; and all occurrences of x in $A(x)$ are bound in $(\forall x.A(x))$).

second-order quantifications: Universal sentences $(\forall F.P)$ where F is a second order variable and P is a formula (all occurrences of F in P are bound in $(\forall F.P)$. Other logical operators (such as the existential quantifier) are to be defined in standard ways.

Since we have introduced second order quantification, we need an axiom and a rule for it.

Q2 $(\forall F.P) \rightarrow P[T/F]$ where $P[T/F]$ is the result of substituting the second order term T for the second order variable F, as long as $P[T/F]$ is well-formed (stratification requires us to make this qualification).

Rule 3 (second order universal generalization): if we can prove P then we can prove $(\forall F.P[F/T])$, where T is a second order constant about which we have made no assumptions and $P[F/T]$ is well-formed.

Observation: The first-order rule of universal generalization can be stated similarly, though the formulation above works. if we can prove P then we can prove $(\forall x.P[x/a])$, where a is a constant about which nothing has been assumed and $P[x/a]$ is the result of substituting x for a, as long as $P[x/a]$ is well-formed. The definition of substitution of course has subtleties caused by variable binding. Substitution of variables for constants can cause ill-formedness because of the stratification discipline.

Second order terms of our language (representing concepts) are of the following forms:

variables: second order variables F, each associated with an integer $\texttt{type}(F)$

constants: second order constants T (to which no type needs to be assigned),

concept terms: concept terms $[A(x)]$, where $A(x)$ is a formula containing at least one free occurrence of the first order variable x and containing no other free variable. The occurrences of x in $[A(x)]$ are bound, and no type needs to be assigned to a concept term.

The whole story about concept terms is contained in this

Rule 4 (concept term application): $[A(x)]t$ may be replaced by $A(t)$ in any context, where t is any first order term, and vice versa.

Concept terms are a novelty here, intended to make instantiation of second order variables clearer. I could similarly reify more general expressions with free variables, obtaining something like the propositional functions of Russell and Whitehead, but this is not needed. Notice that the formation of concept terms is much more restricted than the formation of the first-order abstraction terms.

Finally, we need axioms governing the abstraction terms $\hat{x}A(x)$.

A1 $(\forall x.Fx = Gx) \leftrightarrow \hat{x}Fx = \hat{x}Gx$ (the infamous Axiom V of Frege)

We could readily add a definite description operator as Frege does, but we do not need to do this here.

We can define a membership relation: $t \in u$ abbreviates $(\exists F.u = \hat{x}Fx \wedge Ft)$. Notice that if t and u each contain free variables this is only well-formed if $\texttt{type}(t)+1 = \texttt{type}(u)$. So Russell's paradoxical $\hat{x}(x \not\in x)$ cannot be formed in this system.

But also note that $V = \hat{x}(x = x)$ is well-formed, and $V \in V$ is well-formed and true. But, subtly, this does not mean that we can say $(\exists x.x \in x)$: this is ill-formed.

Note that the defined membership relation satisfies weak extensionality: objects with elements must be abstracts, and Frege's Axiom V entails that abstracts with the same extensions are equal. There may of course be many non-abstracts, all with no elements.

We indicate why stratified comprehension is satisfied. Let $\phi(x)$ be a stratified formula of the language of NFU. Convert the free variables other than x to constants, and the resulting formula ϕ' will be well-formed in the language given here, and $[\phi'(x)]$ will be a concept term. $z \in \hat{x}\phi'(x)$ will be equivalent to $(\exists F.\hat{x}\phi'(x) = \hat{x}Fx \wedge Fz)$: this statement will be true and witnessed by $F = [\phi'(x)]$ iff $\phi'(z)$ is true. The formula $(\exists A.(\forall x.x \in A \leftrightarrow \phi(x)$ is then provable by universal generalization: this is a completely general instance of stratified comprehension.

The interpretation of this system in NFU is very direct. Equality is interpreted by equality. First order variables range over all objects of NFU (atoms and sets). Second order variables range over sets. A proposition Fx translates to $x^* \in F^*$, where x^* translates F and F^* translates F. If the formula ϕ in the language of the theory here translates to ϕ^* in NFU, $\hat{x}\phi(x)$ translates to $\{x \mid \phi^*(x)\}$, and $[\phi(x)]$ also translates to $\{x \mid \phi^*(x)\}$. All axioms of this theory translate to true assertions in NFU, so everything provable in this theory is provable in NFU. It is considerably harder to show, but is true, that every theorem of NFU which is the translation of a sentence of this theory is also a theorem of this theory.

An important point which we of course know is that we have artifically limited the scope of our language: the only primitive predicates that we provide are the binary predicate of equality and monadic second-order constants. We of course know that we can define an ordered pair (and so ordered n-tuple) construction using abstraction, and so represent any desired primitive n-ary predicates we want to add to our system as special second order constants intended to take n-tuples as arguments, but this is an anachronism: this was not known to be the case until Wiener defined the ordered pair in set theory in [?], 1914, and none of Frege, Russell and Whitehead, or Zermelo knew this. It does make our lives simpler, though.

Mathematical validity of this repair is evident. This system can be interpreted in NFU (though its legal formulas are only a subset of the legal formulas of NFU, as for philosophical reasons we do not allow unstratified assertions with free variables to be formed at all). It is not immediately obvious, but follows from work of Marcel Crabbé, that all stratified theorems of NFU can be proved in this system (Crabbé showed in [?] that every stratified theorem has a proof which only mentions stratified formulas).

NFU does not prove Infinity, so an explicit axiom of infinity is needed. Once Infinity is added, this system has the same strength as the theory of types with Infinity. It supports the entire program of the Grundgesetze with ease. This is a fully impredicative mathematical system.

Philosophical adequacy is harder to assess. The question is whether one can justify the rules for typing terms. My view is that there is a coherent philosophical view behind stratification, but that it is one which it would have been difficult to arrive at without seeing in advance that the mathematics works out.

I have talked about this elsewhere. The idea is that while we are working formally in a two sorted theory (first order terms representing objects and second order terms representing first-order concepts), the device of abstraction allows us to represent first order concepts as objects, and iteration of this process allows us to represent concepts of any order by either first or second order terms. The idea is that when we form abstractions we are required to consider any particular object which we have not specified completely (any object represented by a term with free variables) under a single one of these roles (as a concept of a particular order). The syntactical rules for stratification can then be seen to express exactly this restriction. We certainly want the restriction, because otherwise we get paradox. There is an possible *a priori* motivation for the restriction, which is roughly speaking that the representation of concepts by particular objects is not really a feature of either the concept or the object, but an arbitrary device for compacting the levels, and so we should not expect that (for example) non-self-membership is a legitimate feature of a class, because this would be perturbed by a different choice of object to represent the class. This requires considerable exposition, which I have attempted elsewhere.

In the usual formulation of NF or NFU, one allows quantification over unstratified sentences. We do not allow this at all here. The reason is philosophical rather than mathematical, though it does not affect what the theory is able to prove. If one can write $x \in x$ (with x a variable), then there is the Fregean concept of self-membership. We do not want this concept to have any standing at all. We acknowledge that $V \in V$ is true (we can prove it), but we do not allow $x \in x$ to be a general feature of x, and so we do not allow this formula with a variable in it to be formed at all. In this particular case, we can articulate the reasons for this. The issue is that this assertion talks about x in two different roles, as an object and as a first-order concept (or more generally as a concept of order n and as a concept of order $n + 1$, but we stick to the simplest case). Now suppose that x is neither the universe nor the empty set. We could modify the choice of object representing the concept x in such a way as to make $x \in x$ either true or false, without perturbing any essential feature of the concept or of any object. The idea is that we must view the underlying representation of concepts by objects as arbitrary, so that the relation between the concept of order m and the concept of a different order n which happen to be represented by the same object is accidental (in the technical philosophical sense!), and not to be abstracted from.

Incidentally, it is fairly clear that the second order logic is redundant here: all reasoning in the system presented here is supported in first-order NFU. But my intention is to faithfully present Frege's system in a stratified format; that the second order features turn out out be subsumed under first order features of NFU is a discovery about Frege's system, not a criticism. Second order logic over the usual formulation of NFU is of course far more expressive because of the presence of unstratified formulas which correspond to concepts whose extensions cannot be sets.

The reader should compare this with the development given by Nino

Cocchiarella in "Frege's Double-Correlation Thesis and Quine's Set Theories NF and ML" *Journal of Philosophical Logic* Vol. 14, No. 1 (Feb., 1985), pp. 1-39, which gives a similarly motivated full formal development.

I also note that I need to go and read Frege himself carefully!

I acknowledge use of an appendix to Michael Beaney's Frege: Making Sense as a reference for Frege's logical notation and axioms (pp. 283-89). I have looked at Ferreira's "Amending Frege's Grundgesetze", and I have encountered other neo-Fregean systems: this other work generally leads to much weaker systems which do not fully implement Frege's mathematical intentions. This approach clearly fully implements Frege's mathematical intentions (it goes far beyond them); the question is how much violence is done to his philosophy by the technical modification that is applied.