

Frege's Double-Correlation Thesis and Quine's Set Theories **NF** and **ML**

Nino B. Cocchiarella

There are two fundamentally different notions of a class, which, following tradition, we might call the mathematical and the logical notions, respectively. The logical notion is essentially the notion of a class as the extension of a concept, and, following Frege, we will assume that a class in this sense “simply has its being in the concept, not in the objects which belong to it” ([8], p. 183) — regardless of whether or not concepts themselves differ, as Frege assumed, “only so far as their extensions are different” (*ibid.*, p. 118). The mathematical notion of a class, on the other hand, is essentially the notion of a class as composed of its members, i.e., of a class that has its being in the objects that belong to it. This notion of a class, we claim, is none other than the iterative concept of set — or at least that is what it comes to upon analysis. Note that although what accounts for the being of a class under the one notion is not the same as what accounts for the being of a class under the other, nevertheless the axiom of extensionality applies equally to both notions. This means that the axiom of extensionality does not of itself account for the being of a class.¹

Of course the logical notion of a class, especially as developed in Frege's form of logicism, is usually thought to be bankrupt as a result of Russell's paradox. This assessment, however, is erroneous. In particular, in [5] I explained how Frege's view of classes in the logical sense can be reconstructed without paradox by modifying in either of two ways what I there referred to as *Frege's double-correlation thesis*. The two systems that result from these modifications, it turns out, have certain structural similarities with Quine's two set theories **NF** and **ML**, especially when the latter are themselves modified so as to include urelements other than the empty set. This is significant because both **NF** and **ML** are commonly said to “lack a motivation” (cf. [1], p. 219). But this is because as theories of sets in the sense of classes that are composed of their members, which is really the only sense to which Quine is willing to commit himself, both **NF** and **ML** are incompatible with the iterative concept of set. As theories of classes in the logical sense, however, and in particular of the classes that Frege took to be the correlates of concepts, both **NF** and **ML**

¹In an intensional (modal) language, the mathematical notion of a class might well assume a stronger axiom of extensionality, namely, one in which classes that are composed of their members are necessarily identical when they have the same members. Such an axiom would not in general hold for classes in the logical sense, because coextensive concepts are not necessarily coextensive. (It would hold for those classes in the logical sense that are the extensions of “rigid” concepts, i.e., concepts that have the same extension in every possible world.)

can be given a very natural motivation, especially when modified to include urelements. In what follows we will defend this motivation by examining the structural similarities in question.

1 NF and the Iterative Concept of Set

The original motivation for **NF**, according to Quine, was the reconstruction of the theory of simple types as a first-order theory of classes, or what we shall also call a theory of sets in the sense of classes that are composed of their members. The theory of simple types, as a theory of classes, was already itself a reconstruction by Ramsey of Russell's earlier theory of ramified types as a theory of intensional concepts (or what Russell called propositional functions). The latter reconstruction was implicit in the interpretation Russell gave to his axiom of reducibility (cf. [21], p. 58 and p. 166), but whereas for Ramsey classes were real and had their being in the objects that belonged to them, Russell chose not to assume that classes had any being at all but rather that all our talk of classes was to be explained in terms of concepts. Quine, of course, eschews all intensional entities and saw in the theory of simple types only a theory of sets in the sense of classes that are composed of their members. (For Quine, apparently, the axiom of extensionality provides a full account of the being of a class.) The real ontological import of the theory of simple types, according to Quine, is a stratified comprehension principle for sets which it trivially effects through its overly restrictive grammatical constraints. If we delete these grammatical constraints as being improperly retained through their historical connection with the intensional theory of ramified types, then what remains is none other than the stratified comprehension principle for sets given in **NF**.

Quine also assumed, but only as a "harmless" simplifying gesture, that the axiom of extensionality was to be applied to all objects whatsoever. It follows of course that there is then but one urelement, namely, the empty set. Objects that are not sets, according to Quine, can be simply identified with their singletons, though of course no commitment need actually be made in **NF** that there are any such objects at all. Stated as an applied first-order theory (with identity), the system **NF** is really quite elegant in its simplicity.

But all is not well with **NF** as a theory of sets, i.e., as a theory of classes that are composed of their members. For the notion of a set as a class which is composed of its members, i.e., of a class that has its being in the objects that belong to it, is none other, we maintain, than the iterative concept of set. And **NF**, apparently, is incompatible with any theory that can be said to represent the iterative concept of set. Thus, the standard first-order set theory that is taken as a paradigm of systems that represent the iterative concept of set is **ZF** (Zermelo-Fraenkel set theory²), and the incompatibility between **ZF** and **NF** is easy to demonstrate. In particular, in **ZF** there neither is nor can be a universal

² We ignore the question here of whether or in what sense the axiom of replacement accords with the iterative concept of set (cf. [1], section IV). We could just as well speak of Zermelo set theory instead.

set, and therefore no set has an absolute complement. In addition, in **ZF** no set can be a member of itself. In **NF**, on the other hand, there is a universal set, which as such is therefore a member of itself; and of course every set has an absolute complement.

The guiding principle of **ZF**, incidentally, is what Russell once called the limitation of size doctrine, i.e., the doctrine that sets are not to get too big too fast (cf. [20], p. 152). According to this doctrine, we are “to admit only those instances of the axiom schema of comprehension which assert the existence of sets which are not too ‘big’ compared to sets which we already have” ([10], p. 32). Of course, it is just this doctrine that is realized by the iterative concept of set, since according to the latter sets are formed at stages by iterating the operation: $X \rightarrow X \cup \mathcal{P}(x)$, where $\mathcal{P}(x)$ is the power set of X . It is by identifying the being of a set with its “formation” in this manner, incidentally, that we are able to say that the iterative concept of set is none other than the analysans of the mathematical notion of a set as a class that is composed of its members, i.e., of a class that has its being in the objects which belong to it.

Of course, the stages at which sets are formed are themselves well-ordered, and although the strongest form of the iterative concept of set assumes that for any given ordinal α there is an α th stage (cf. [25], Chapter VI), a weaker form assumes only that there is a stage corresponding to some limit ordinal (cf. [1], p. 221). But even this is not really essential to the concept (cf. [15], p. 98), and in fact the theory of sets (sometimes also called *general set theory*) obtained from **ZF** by omitting the axiom of infinity might be said to represent the core form of the iterative concept. This is important, since if we exclude infinite stages and ignore the cumulative nature of the operation in question, then what we get by representing the stages by different types is essentially the theory of simple types as a theory of classes that are composed of their members. As a theory of classes, in other words, the theory simple types represents a restricted form of the iterative concept of set, and in fact it represents a form which has its origin in Schröder’s *Algebra der Logik*.

Schröder, as Alonzo Church has observed, anticipated the theory of types (as a theory of classes) when he took “the universal class 1 which appears in his algebra, not as an absolute universal class, but as composed of all the elements of a certain domain fixed in advance” [2], p. 150). Once such a quasi-universal class or “manifold” was given, a second may then be obtained (to which the algebra was to be applied in turn) “by taking the subsets of the first to be the individuals of the second” (*ibid.*); and by continuing in this way a “hierarchy of *reine Mannigfaltigkeiten* may be extended to infinity” (*ibid.*). The important restriction Schröder imposed on this hierarchy was that no subset of the domain of “individuals” considered at any one stage of the hierarchy was to be among the individuals of that stage, and that consequently the laws of logic, as the laws of this algebra, were to be restricted in any given application only to the “individuals” of the stage in question. Russell’s first theory of types in Appendix B of his *Principles of Mathematics* was essentially this theory of classes, incidentally, though to what extent Russell was himself aware of this fact is not clear (cf. [5], Section 8). That is, it was classes and not propositional

functions to which Russell assigned types in that appendix; and the classes in question, moreover, were “classes as many”; i.e., classes that essentially were nothing but their members. (It was not until after publication of the *Principles* that Russell adopted his famous “no classes” theory.)

Now as a restricted form of the iterative concept of set, the theory of simple types as a theory of classes also realizes a restricted form of the limitation of size doctrine; and, indeed, it is for precisely this reason that an axiom of infinity to the effect that there are infinitely many urelements is needed in type theory. Also, just as the axiom of choice is neither provable nor disprovable in **ZF** (and therefore not really involved in the iterative concept of set), it is also neither provable nor disprovable in the theory of simple types as a theory of classes. Yet, in **NF**, which is supposed to be a first-order reconstruction of the theory of simple types as a theory of classes, the axiom of choice is disprovable, and therefore, because the axiom is provable for finite sets, it follows that some set in **NF** is infinite (cf. [22]). In **NF**, moreover, because it contains the Boolean calculus of classes, there is a complete symmetry between the existence or being of sets that are small and the existence or being of sets that are large: that is, the doctrine of the limitation of size is completely rejected in **NF**. It is no wonder then that as a theory of sets, i.e., as a theory of classes that are composed of their members, **NF** is said to “lack a motivation”.

The alternative to this conclusion, of course, is that **NF** is really not a theory of sets in the sense of classes that are composed of their members. That is, if it has any philosophically coherent motivation at all, then it is as a theory of classes in the logical sense, i.e., in the sense in which classes have their being in the concepts whose extensions they are, and not as a theory of classes that have their being in the objects that belong to them. Such an alternative, it is clear, requires a background theory of concepts, and in particular a theory of predicable concepts whose extensions are classes in the logical sense. That is, such an alternative presupposes a superseding theory of predication in which predicates, despite their not being singular terms, can be said to have referential significance, and therefore one in which quantification with respect to predicate positions can be said to be significant as well. This conclusion, needless to say, would not be welcome to anyone, such as Quine himself, who rejects the idea that predicate quantifiers can have any referential significance at all. For Frege, however, the notion that predicate quantifiers have a referential significance even though predicates are not themselves singular terms is not only acceptable but is in fact forced upon us by the very nature of language and thought (cf. [5], Sections 1 and 2).

2 The Theory of Simple Types and Frege’s Double-Correlation Thesis

It is sometimes claimed, as Church observed (long before the view became fashionable), that Frege’s *Stufen* or levels of concepts “constitute an anticipation

of the simple theory of types" ([2], p. 151). Church himself rejects this claim, however, and he is right in doing so if by the theory of simple types we mean either a theory of classes in the sense already described or a theory of concepts all of which can have classes as their extensions in the same sense in which first-level concepts have classes as their extensions. Of course, in some sense Frege's theory of different levels of concepts does amount to a theory of types. But still, unlike the situation in Russell's theory of types where all propositional functions have extensions in the same sense in which the first-order propositional functions that apply to concrete individuals do, in Frege's theory only first-level concepts have or even can have extensions. To be sure, according to Frege, every concept of a given level *falls within* some concept of the next higher-level; but still, no concept in Frege's hierarchy can be said to *fall under* a higher-level concept the way that objects fall under first-level concepts.

Nor is this merely a verbal distinction either! E.g., second-level concepts correspond not to predicates the way that first-level concepts do, but to variable-binding operators that apply to wffs (well-formed formulas) and result in wffs when so applied (such as the existential quantifier when affixed to an individual variable). (Equivalently, second-level concepts correspond to wffs with free predicate variables that may be used in a third-level comprehension principle to specify such a variable binding-operator, such as the wff $\forall x[F(x) \rightarrow G(x)]$, which specifies the second-level relation of subordination between one first-level concept and another (cf. [3].) Unlike predicates, these operators can be commuted and iterated, as well as occur within the scope of one another. And of course the same observation applies to operators for third and higher-level concepts, except that these operators can also be commuted with those for second-level concepts, and therefore they can occur within the scope of the latter as well. Is this important? Indeed it is, for contrary to the view of the theory of types as a restricted form of the iterative concept of set, what it seemed to suggest to Frege is that there are no more third- and higher-level concepts than there are second-level concepts; and, apparently, this is why he saw no point in explicitly introducing notation for third- and higher-level concepts into his ideography (cf. [9], p. 41).³

Frege also saw no point in introducing quantifiers regarding second-level concepts; and the reason why is that he also assumed that there are no more second-level concepts than there are first-level concepts. That is, according to Frege, and this is what we mean by his double-correlation thesis, all second-level concepts can be correlated one-to-one with certain special first-level concepts, which in turn can be correlated one-to-one with special objects called concept-correlates. These concept-correlates are none other than Frege's value-ranges

³One can of course *model* variable-binding operators in terms of sets (as is done in [3]). But in that case, at least if the operators are understood to stand for Frege's higher-level concepts, the modeling must be restricted to Henkin's *general* models for type theory (cf. [13]). To insist that general models are not the *real* or "primary" semantics for variable-binding operators is to confuse Frege's theory of types (if one insists on calling it that) with Russell's — or what essentially comes to the same thing given Frege's extensional view of concepts, it is to confuse Frege's hierarchy of concepts with a restricted form of the iterative concept of set.

(*Wertverläufe*), which in the monadic case Frege also called classes (*Begriffsumfängen*). Thus quantifiers regarding second-level concepts are unnecessary, according to Frege, because second-level concepts “can be represented in a certain manner” by first-level concepts whereby the concepts “that appear as arguments of the former are represented by their value-ranges” ([7], Section 25).⁴ That is, in symbols (in the case of unary concepts):

$$\forall Q \exists F \forall G [QxG(x) \leftrightarrow F(\acute{x}G(x))].$$

In the relational case, incidentally, Frege’s basic law V, *namely*,

$$\acute{x}F(x) = \acute{x}G(x) \leftrightarrow \forall x[F(x) \leftrightarrow G(x)],$$

is none other than a special instance of his double-correlation thesis. For what is indicated on the right side of this law is none other than the second-level relation of material equivalence or mutual subordination of two first-level concepts; and on Frege’s extensional view of concepts (as functions from objects to truth-values) such an equivalence amounts in effect to their “identity”. That is, Frege’s basic law V amounts to correlating the first-level relation of identity with his second-level relation of mutual subordination. Such a correlation is needed, Frege observed, because “to construe mutual subordination simply as equality is forbidden by the basic difference between first- and second-level relations. Concepts cannot stand in a first-level relation. That wouldn’t be false, it would be nonsense. Only in the case of objects can there be any question of equality (identity). And so *the said transformation* [from mutual subordination to identity] *can only occur by concepts being correlated with the same object*. It is all, so to speak, moved lawn a level” [8], p. 182, italics added).

Now just as first-level concepts can be “moved down a level” by being correlated with their extensions, so too, by Frege’s double-correlation thesis, second-level concepts can be moved down a level by being correlated with certain special first-level concepts, which in turn of course are correlated with their own extensions. But then, third-level concepts can be similarly “moved down a level” by being correlated with second-level concepts, and therefore by the product of these correlations, third-level concepts can in effect be correlated with first-level concepts, which again are correlated with their own extensions. Similarly, fourth-level concepts can be correlated with third-level concepts and therefore with first-level concepts as well. In general, in other words, all concepts of whatever level can in effect be correlated with first-level concepts, and these in turn can be correlated with their extensions. Frege’s entire hierarchy of concepts, can be collapsed into the universe of first-level concepts, which in turn can be correlated with the classes that are their extensions. In this regard, the theory of classes that is contained in Frege’s theory of types of concepts is radically

⁴Frege’s thesis is stated for all second-level *functions*, not just concepts, i.e., not just for those functions that have truth-values as their values. For convenience (and otherwise for reasons discussed in [5], Section 2), we will ignore functions other than concepts. Also, although we will in general restrict ourselves to the monadic case throughout, we will also sometimes refer to relations as concepts.

different from the theory of simple types as itself a theory of classes, i.e., it is radically different from the theory of simple types as a restricted form of the iterative concept of set.

Of course, Frege's theory, as originally described, is subject to Russell's paradox. But this result has nothing to do with the difference between Frege's hierarchy of concepts and the theory of simple types as a restricted form of the iterative concept of set. Indeed, by modifying in either of two ways Frege's double correlation thesis, we can reconstruct Frege's theory so that the resulting framework is provably consistent if the theory of simple types as a restricted form of the iterative concept of set is consistent. The two alternative systems resulting by these reconstructions, it turns out, are similar in structure to Quine's set theories **NF** and **ML**, and of course that suggests re-evaluating the latter in the light of our two modifications of Frege's double correlation thesis.

3 Frege's Logicism as a Second-Order Predicate Logic with Nominalized Predicates

Frege's original *Begriffsschrift*, as is well known, amounts to the first formulation ever given of standard second-order predicate logic. It is clear of course that Frege viewed his system as providing a logical analysis of the predicates that occur in natural language. In the *Grundgesetze*, Frege supplemented this predicate logic with his smooth=breathing abstraction operator that when applied to wffs generated complex singular terms. These singular terms were of course taken as denoting value-ranges, which for reasons already indicated, Frege also called concept-correlates. What we have not yet noted, however, is that these concept-correlates are also none other than the objects that Frege took the nominalized predicates of natural language to denote (cf. 151, Sections 4 and 5). In this regard, we maintain, Frege intended that his overall system was to provide a logical analysis not only of the predicates that occur in natural language but also of the predicate nominalizations that occur therein as well. For this, among other reasons, it was suggested in [5] that the confusion sometimes made of Frege's theory of value-ranges with a second-order set theory might best be obviated by directly describing his theory of value-ranges as a theory of concept-correlates in the sense of the objects that are denoted by nominalized predicates. In other words, Frege's form of logicism, we maintain, can best be described as a second=order predicate logic with nominalized predicates. That, in any case, is the approach we shall adopt here.

In doing so, however, we will for convenience allow ourselves certain liberties in the way we will represent predicates and their nominalizations. In particular, we will represent the unsaturated nature of predicates (at least when they occur as such) by the requirement that they be accompanied by a pair of parentheses (and commas as well in the case of a relational predicate). We will then represent the nominalization of a predicate by the deletion of these parentheses (and the accompanying commas, if any). Thus, whereas $F(x)$ and $R(x, y)$ are wffs in

which F and R occur as predicates, $G(F)$ and $G(R)$ are wffs in which F and R occur as singular terms. In $F(F)$ and $R(F, R)$, of course, F and R occur both as predicates and as singular terms (though no single occurrence can be both as a predicate and as a singular term). The predicate expressions that occur in the argument or subject positions of other predicates, or of themselves as well, needless to say, are intended as representing the nominalized predicates that occur in natural language. Note that adding such suffixes as ‘-ity’, ‘-ness’ or ‘-hood’ to nominalized occurrences of predicates would be completely superfluous here, because such occurrences are already formally identified as subject position occurrences; and of course the same observation applies to such related phrases as ‘the concept F ’ or ‘being an F ’. Such phrases and suffixes are important in English syntax, no doubt, because they serve to mark derived nominal expressions in the surface grammar of English: but it is sheer sophistry to insist that such surface grammatical features of English either must or should occur in our logical forms, as though a logical error were being committed otherwise.

In addition, we will also use Church’s λ -operator for the formation of complex predicates. (This is in keeping with most of the higher-order predicate logics formulated in the recent literature.) Thus, where $\varphi(x)$ is a wff, we also have $[\lambda x\varphi(x)]()$ as a complex predicate (note the accompanying pair of parentheses) and $[\lambda x\varphi(x)]$ as a complex singular term. For convenience, however, we will usually drop the accompanying parentheses (and commas) when referring to predicates (as unsaturated expressions), though always, whenever a predicate occurs in a wff as a predicate, it will have an accompanying pair of parentheses (and commas as well if it is a relational predicate).

Finally, in describing our logical grammar we will for convenience identify the different types of meaningful expressions by associating them with different natural numbers. Thus we will understand 0 to represent the type of all *singular terms*, 1 the type of all wffs (propositional forms), and $n + 1$, for all $n > 0$, as the type of all n -place predicate expressions. (We will ignore the introduction of predicate and individual constants in the present paper.) We assume of course the availability of denumerably many individual variables, and, for each natural n denumerably many n -place predicate variables. (We identify propositional variables with the 0-place predicate variables.) We take ‘ x ’, ‘ y ’, ‘ z ’ and ‘ w ’, with or without numerical subscripts, to refer (in the metalanguage) to individual variables, and ‘ F^n ’, ‘ G^n ’, ‘ H^n ’, ‘ R^n ’ to similarly refer to n -place predicate variables. (We will usually drop the superscript when the context makes clear the degree of a predicate variable.) We will also use ‘ u ’ to refer to variables in general, and ‘ a ’ and ‘ b ’ to refer to singular terms in general. We take \rightarrow , \neg , $=$, \forall , λ as primitive logical constants and assume the others to be defined (as abbreviations of the metalanguage) in the usual way. For each natural number n , accordingly, we recursively define the meaningful expressions of type n , in symbols, \mathbf{ME}_n , as follows:

1. every individual variable (or constant) is in \mathbf{ME}_0 , and every n -place predicate variable (or constant) is in both \mathbf{ME}_{n+1} and \mathbf{ME}_0 ;

2. if $a, b \in \mathbf{ME}_0$, then $(a = b) \in \mathbf{ME}_1$;
3. if $\pi \in \mathbf{ME}_{n+1}$, and $a_1, \dots, a_n \in \mathbf{ME}_0$, then $\pi(a_1, \dots, a_n) \in \mathbf{ME}_1$;
4. if $\varphi \in \mathbf{ME}_1$, and x_1, \dots, x_n are pairwise distinct individual variables, then $[\lambda x_1 \dots x_n \varphi] \in \mathbf{ME}_n$;
5. if $\varphi \in \mathbf{ME}_1$, then $\neg \varphi \in \mathbf{ME}_1$;
6. if $\varphi, \psi \in \mathbf{ME}_1$, then $(\varphi \rightarrow \psi) \in \mathbf{ME}_1$;
7. if $\varphi \in \mathbf{ME}_1$, and a is an individual or a predicate variable, then $\forall a \varphi \in \mathbf{ME}_1$;
8. if $\varphi \in \mathbf{ME}_1$, then $[\lambda \varphi] \in \mathbf{ME}_0$; and
9. if $n > 1$, then $\mathbf{ME}_n \subseteq \mathbf{ME}_0$.

Singular terms, which we shall also refer to simply as *terms*, are understood to be all the members of \mathbf{ME}_0 ; and for $n > 0$, we understand the members of \mathbf{ME}_{n+1} to be n -place predicate expressions. Wffs or propositional forms are of course all the members of \mathbf{ME}_1 . Note that whereas by clause (9) every predicate expression is a term, not every wff is a term. We differ in this regard from what Frege would allow; but the difference is unimportant since by clause (4), where $n = 0$, $[\lambda \varphi]$ is a wff iff φ is a wff, and by clause (8) $[\lambda \varphi]$ is a term. In other words, besides 0-place predicate variables (and constants), wffs are terms only when prefixed by the λ -operator. Thus, in general, we may read ‘ $[\lambda \varphi]$ ’ as ‘that φ ’ when it occurs in a wff as a term, i.e., when it occurs in one of the argument or subject positions of a predicate expression.

If we take this logical grammar as a counterpart to Frege’s own original ideography, then, ignoring Russell’s paradox for the present, the axioms of standard second-order predicate logic when applied to this grammar can be taken as the counterparts to Frege’s basic laws of logic. The following axiom set will serve us for this purpose. Of course we include an axiom for λ -conversion, and an axiom that identifies what a nominalized predicate variable (or constant) denotes with what its corresponding nominalized λ -abstract denotes. Also, we add a *-label to all theses that are understood to include wffs having nominalized predicates as singular terms among their instances. Bondage and freedom are understood defined in the usual way.

- (A0*) all tautologous wffs,
- (A1*) $\forall u(\varphi \rightarrow \psi) \rightarrow (\forall u \varphi \rightarrow \forall u \psi)$,
- (A2*) $\varphi \rightarrow \forall u \varphi$, where u is an individual or predicate variable not occurring free in φ ,
- (A3*) $\exists x(a = x)$, where a is a singular term in which x does not occur free,
- (LL*) $(a = b) \rightarrow (\varphi \leftrightarrow \psi)$, where a, b are singular terms and ψ comes from φ by replacing one or more free occurrences of b by free occurrences of a ,

- (CP*) $\exists F^n \forall x_1 \dots \forall x_n [F(x_1, \dots, x_n) \leftrightarrow \varphi]$, where F^n does not occur free in φ and x_1, \dots, x_n are distinct individual variables,
- (λ -Conv*) $[\lambda x_n \dots x_n \varphi](a_1, \dots, a_n) \leftrightarrow \varphi(a_1/x_1, \dots, a_n/x_n)$, where a_1, \dots, a_n are singular terms and each a_i is free for x_i in φ ,
- (Id $_{\lambda}^*$) $[\lambda x_n \dots x_n R(x_1, \dots, x_n)] = R$, where R is an n -place predicate variable (or constant).

As inference rules, we shall assume modus ponens (MP) and universal generalization (UG) as applied to an individual or a predicate variable. Note that by Leibniz's law, i.e., (LL*), and (A0*)–(A3*), the principle of universal instantiation of a singular term for an individual variable is provable:

$$(UI_1^*) \quad \forall x \varphi \rightarrow \varphi(a/x).$$

This is essentially Frege's basic law (IIa). Frege's basic law (IIb) is similarly derivable from the second-order counterpart of (LL*) and by using (CP*) instead of (A3*). Note also, incidentally, that (LL*) and (λ -Conv*) together yield the following generalized form of Frege's basic law (Vb):

$$[\lambda x_1 \dots x_n \varphi] = [\lambda x_1 \dots x_n \psi] \rightarrow \forall x_1 \dots \forall x_n (\varphi \leftrightarrow \psi).$$

Of course, this is only one direction of Frege's basic law V. The other direction, i.e., Frege's basic law (Va), is the following *principle of extensionality*, which we will consider separately from the other axioms:

$$\forall x_1 \dots \forall x_n (\varphi \leftrightarrow \psi) \rightarrow [\lambda x_1 \dots x_n \varphi] = [\lambda x_1 \dots x_n \psi]. \quad (Ext^*)$$

Finally, although we have not included variables for variable-binding operators representing second-level concepts, note that were such variables available, we could represent Frege's double-correlation thesis as follows:

$$\forall Q \exists F \forall G [\mathcal{Q}xG(x) \leftrightarrow F(G)].$$

Now since $\mathcal{Q}xG(x)$ can also be taken as an abbreviation of a wff that contains G as a free 1-place predicate variable, i.e., a wff that could be used to specify a second-level concept \mathcal{Q} in a third-level comprehension principle, then, as Frege himself noted, we can ignore explicitly introducing quantifiers for second-level concepts and capture the correlation in question as an instance of (CP*) instead. For by (CP*),

$$\exists F \forall x (F(x) \leftrightarrow \exists G [x = G \wedge \mathcal{Q}xG(x)]),$$

and therefore by (UI*) and other obvious transformations

$$\exists F \forall G [F(G) \leftrightarrow \mathcal{Q}xG(x)].$$

In other words, every instance of Frege's double-correlation thesis can be captured as an instance of (CP*).

It is clear, accordingly, that together with (Ext*), the above system of axioms and rules amounts to a counterpart of Frege's own original system. Of course,

like Frege's original system, the above system, with or without (*Ext**), is subject to Russell's paradox and therefore is inconsistent. That is, by the following instance of (CP^*),

$$\exists F \forall x (F(x) \leftrightarrow \exists G [x = G \wedge \neg G(x)]),$$

a contradiction is provable in the system. Nevertheless, the framework is repairable, and by a simple modification of Frege's double-correlation thesis as well; i.e., by a simple modification of (CP^*) as the representation of Frege's thesis in the present context. Before turning to this modification, however, let us note that (CP^*) is a consequence of a still simpler form of comprehension principle: namely,

$$\exists F^n (F = [\lambda x_1 \dots x_n \varphi]), \quad (\text{CP}^*_\lambda)$$

where F^n does not occur free in φ . Since a restricted form of (CP^*_λ) will yield only a restricted form of (CP^*), we will apply our modification to (CP^*_λ) instead. Note, however, that since (CP^*_λ) posits the existence of a concept corresponding to any given λ -abstract, then the restriction in question really amounts to a restriction as to which λ -abstracts are to be regarded as well-formed in the first place.

4 The Theory of Homogeneous Simple Types as a Second-Order Predicate Logic

Now the point of our first modification of Frege's double-correlation thesis is that if all higher-level concepts can really be correlated with first-level concepts, which in turn are correlated with their concept-correlates, then the conditions for specifying the first-level concepts in question must be stratified in a way corresponding to the stratification of the higher-level concepts to which these first-level concepts and their concept-correlates correspond. The comprehension principle (CP^*_λ), in outer words, must be restricted in a way that corresponds to the stratification of the unsaturated concepts of Frege's hierarchy.

For reasons indicated below, however, we will actually need a more stringent form of stratification than Frege allows. In particular, we will be required to assume, on pain otherwise of generating Russell's paradox after all, that higher-level relations are only homogeneously stratified. This in fact is the only modification of Frege's original form of logicism that we will assume in our first reconstruction of Frege's logic.

Returning to the logical grammar of Section 3, accordingly, let us say that a wff or λ -abstract φ of that grammar is *homogeneously stratified* iff there is an assignment t of natural numbers to the set of terms occurring in φ (including φ itself if φ is a λ -abstract) such that (1) for all terms a and b , if $(a = b)$ occurs in φ , then $t(a) = t(b)$; (2) for all $n \geq 1$, all n -place predicate expressions π and all terms a_1, \dots, a_n , if $\pi(a_1, \dots, a_n)$ is a wff occurring in φ , then (i) $t(a_j) = t(a_k)$, for $1 \leq j, k \leq n$, and (ii) $t(\pi) = t(a_1) + 1$; (3) for all natural numbers m , all individual variables x_1, \dots, x_m , and all wffs ψ , if $[\lambda x_1 \dots x_m \psi]$ occurs in φ , then

(iii) $t(x_j) = t(x_k)$, for $1 \leq j, k \leq m$, and (iv) $t([\lambda x_1 \dots x_m \psi]) = t(x_1) + 1$; and (4) for all formulas χ , if $[\lambda \chi]$ occurs in φ and a_1, \dots, a_k are all of the terms or predicates occurring in χ , then $t([\chi]) \geq \max[t(a_1), \dots, t(a_k)]$.

If clauses (i) and (iii) are dropped and clauses (ii) and (iv) are replaced by the weaker requirement that $t(\pi) = 1 + \max[t(a_1), \dots, t(a_m)]$ and $t([\lambda x_1 \dots x_m \psi]) = 1 + \max[t(x_1), \dots, t(x_m)]$, then we will say that φ is *heterogeneously stratified*. Of course, by definition, every homogeneously stratified expression is heterogeneously stratified; and although not all heterogeneously stratified expressions are homogeneously stratified, nevertheless, a monadic expression, i.e., one in which every predicate is a 1-place predicate, is homogeneously stratified if and only if it is heterogeneously stratified. For this reason, homogeneously stratified monadic wffs will be referred to hereafter simply as stratified wffs.

Now it is clear that Frege's hierarchy of concepts is heterogeneously stratified and not just homogeneously stratified; i.e., some of Frege's higher-level relations are unequal-leveled relations. Thus, e.g., the second-level relation of an object to a concept under which that objects falls is said by Frege to be an unequal second-level relation because it has as arguments both a saturated object and an unsaturated first-level concept (cf. [7], Section 22). The relevant modification of his double-correlation thesis, accordingly, would seem to be that we are to admit as well-formed only those λ -abstracts that are heterogeneously stratified. This would certainly exclude the version of Russell's paradox described above, because the complex predicate in question, namely $[\lambda x \exists G(x = G \wedge \neg G(x))]$, is not heterogeneously stratified.

Unfortunately, this proposal does not also exclude other forms of Russell's paradox generated in terms of unequal leveled concepts. E.g., since Frege's unequal-leveled relation of subsumption under a concept is specified by a heterogeneously stratified λ -abstract, namely, $[\lambda xy \exists G(x = G \wedge G(y))]$, then being impredicable with respect to this relation is also specifiable in terms of such a λ -abstract, namely $[\lambda z \neg [\lambda xv \exists G(x = G \wedge G(yv))](z, z)]$. But then Russell's paradox of the concept that is subsumed under itself if and only if it is not subsumed under itself is derivable after all. It follows, accordingly, that if Frege's double-correlation thesis is to apply to all higher-level unsaturated relations, then we must assume that none of these relations are inhomogeneously stratified, i.e., that there are no unequal higher-level relations. In particular, we must not assume that there can be such an unequal second-level relation as the subsumption of an object under a concept; for it is fundamental to Russell's paradox that predication cannot stand for a relation between an object and a concept-correlate.

On the modification in question, accordingly, a λ -abstract is considered as well-formed if and only if it is homogeneously stratified; and, in general, let us say that a meaningful expression of type n is a meaningful expression of λHST^* (where λHST^* is the $*$ -version of the theory of homogeneous simple types) if and only if every λ -abstract occurring in that expression is homogeneously stratified. (Note that this does not require that wffs themselves be homogeneously stratified; in particular, $F(F)$ and $\neg F(F)$ are both wffs of λHST^* , and if $[\lambda x \varphi]$ is homogeneously stratified, then both $[\lambda x \varphi]([\lambda x \varphi])$ and its negation are wffs

of λHST^* as well.) By an axiom of λHST^* , moreover, let us understand any wff of λHST^* that is an instance of $(A0^*)$ – $(A3^*)$, (LL^*) , (CP_λ^*) , $(\lambda\text{-Conv}^*)$ and (Id_λ^*) , i.e., of the axioms already described in Section 3. For convenience, we will refer to the restriction of (CP_λ^*) that is effected hereby as *the homogeneously stratified comprehension principle*, or more simply as $(HSCP_\lambda^*)$. Our proposal or first alternative reconstruction of Frege’s form of logicism, accordingly, is that it be restricted to the meaningful expressions of λHST^* and that it be represented by the axiom system $\lambda HST^* + (Ext^*)$. (We assume, as before, that (MP) and (UG) are the only rules of inference.⁵)

We need not follow Frege in construing concepts extensionally, incidentally. E.g., we could also take \Box as a primitive logical constant and add to the axioms of λHST^* the axiom schemas of the $S5$ modal propositional logic (and the rule of modal generalization as well). In that case, instead of (Ext^*) we might assume the corresponding principle of intensionality:

$$\Box \forall x_1 \dots \forall x_n (\varphi \leftrightarrow \psi) \rightarrow [\lambda x_1 \dots x_n \varphi] = [\lambda x_1 \dots x_n \psi] \quad (\Box Ext^*)$$

If we refer to this extension of λHST^* as $\Box \lambda HST^*$, then Russell’s idea of reducing all talk of classes to talk of concepts can be captured in terms of the following notion of *rigidity*:⁶

$$Rigid =_{df} [\lambda x \exists F (x = F \wedge \forall y [\Box F(y) \vee \Box \neg F(y)])].$$

This suffices because a class, on this analysis, is simply the correlate of a rigid intensional concept:

$$Cls =_{df} [\lambda x \exists F (x = F \wedge Rigid(F))].$$

The “law of logic” that is needed in that case is the following *principle of rigidity*:

$$\forall F \exists G (Cls(G) \wedge \forall x [F(x) \leftrightarrow G(x)]). \quad (PR)$$

It is clear, of course, that by translating \Box as double negation, $\lambda HST^* + (\Box Ext^*) + (PR)$ reduces to just $\lambda HST^* + (Ext^*)$, and therefore since the latter is translatable into the former it follows that the two systems are equiconsistent.⁷

It should be noted in this context, incidentally, that in $\Box \lambda HST^*$ not all concepts are rigid, and therefore not all concept-correlates are “classes”. They

⁵The rewrite law,

$$[\lambda x_1 \dots x_n \varphi] = [\lambda y_1 \dots y_n \varphi(y_1/x_1, \dots, y_n/x_n)],$$

where no y_i occurs in φ , is derivable in λHST^* on the basis of (Ext^*) — or its intensional counterpart $(\Box Ext^*)$ described later in this paper. If neither of these principle is assumed, then the rewrite law must be taken as an additional axiom schema of λHST^* , or of $\Box \lambda HST^*$ as described later.

⁶This notion occurs in Montague [16], p. 132. It also occurs in [11], p. 77, where the principle of extensional comprehension is the type-theoretical counterpart of the principle of rigidity described below in the text.

⁷Although this last claim is correct as it stands, it would nevertheless be more natural to extend both the notion and the principle of rigidity to relations as well (cf. [5], section 13).

might, of course, be none other than the intensional concepts themselves—that is, if, along with Russell, we assume that the latter do not have an unsaturated nature after all. In that case, $\lambda HST^* + (\Box Ext^*) + (PR)$ might well be taken as representing a form of logicism that Russell himself seemed to be defending in his 1902–1904 correspondence with Frege. (Cf. [5] for a defense of this claim.) There are other options here as well, however, but we will not discuss them in the present paper (cf. [5], Section 16). In addition, we might even forego assuming either (Ext^*) or $(\Box Ext^*)$, though in that case the idea of classes as concept-correlates (whether of all concepts or of only rigid concepts) must then be discarded.⁸

5 Quine’s Thesis and the Similarity of NF with $\lambda HST^* + (\Box Ext^*) + (Q^*)$

Although motivated in terms of different notions of a class, the similarity between Quine’s “set” theory **NF** and $\lambda HST^* + (Ext^*)$ as the result of our first modification of Frege’s double-correlation thesis should by now be obvious. We will examine the details of this similarity by restricting ourselves to monadic $\lambda HST^* + (Ext^*)$. (The restriction is inessential, because monadic $\lambda HST^* + (Ext^*)$ is equiconsistent with the full system — cf. [6], Chapter 4, Section 8.)

First, let us note that because urelements other than the empty class are represented in **NF** by their singletons or unit classes, then we will have to make a comparable assumption for $\lambda HST^* + (Ext^*)$. Although the assumption we give is actually simpler than Quine’s, we will nevertheless refer to it as Quine’s thesis. It is the assumption that every object is a concept-correlate:

$$\forall x \exists F (x = F). \quad (Q^*)$$

Our initial result, accordingly, is that by defining membership as follows:

$$x \in y =_{df} \exists F [y = F \wedge F(x)],$$

every theorem of **NF** is a theorem of $\lambda HST^* + (Ext^*) + (Q^*)$; i.e., **NF** is contained in $\lambda HST^* + (Ext^*) + (Q^*)$ and therefore **NF**-“sets” can be construed as Fregean concept-correlates. (We will hereafter assume this analysis of membership when referring to classes as concept-correlates, i.e., as the objects denoted by nominalized predicates. Because stratification in **NF** coincides under this analysis with stratification in λHST^* , incidentally, we will not bother redefining it for first-order wffs with \in as a primitive 2-place predicate constant.)

By way of proof of this result, note that by definition,

$$\vdash_{\lambda HST^*} x \in F \leftrightarrow F(x).$$

⁸See [6], Chapter 6, Section 5 for a semantics of λHST^* that does not validate either (Ext^*) or $(\Box Ext^*)$.

and therefore since by applying (UG) to (A3*),

$$\vdash_{\lambda HST^*} \forall F \exists y (F = y),$$

then by ($HSCP_\lambda^*$),

$$\vdash_{\lambda HST^*} \exists y \forall x (x \in y \leftrightarrow \varphi),$$

where φ is a stratified monadic wff in which y does not occur free. It follows, accordingly, that every instance of the stratified comprehension principle for **NF**-“sets” (when membership is analyzed as above) is a theorem of λHST^* . In regard to the extensionality axiom,

$$\forall x \forall y (\forall z [z \in x \leftrightarrow z \in y] \rightarrow x = y), \quad (Ext)$$

we note that in fact

$$(Ext^*) \vdash_{\lambda HST^*} (Q^*) \leftrightarrow (Ext),$$

and therefore that Quine’s thesis (Q^*) is provably equivalent to the axiom of extensionality for **NF**-“sets” (which is another reason for calling it Quine’s thesis). Because the proof proceeds by obvious transformations, we will not go into the details here (especially since they can already be found in [4], p. 511).

Besides **NF** being contained in $\lambda HST^* + (Ext^*) + (Q^*)$, it is clear that by (mis)construing predicates as singular terms and interpreting predication as membership, every theorem of $\lambda HST^* + (Ext^*) + (Q^*)$ will be translated into a theorem of **NF**. In other words, $\lambda HST^* + (Ext^*) + (Q^*)$ and each of its subsystems is consistent relative to **NF**.

In defining the translation function in question, incidentally, we will make use of the abbreviatory abstraction notation Quine introduced in his discussion of **NF** and **ML**. The abbreviations in question, it should be noted, are actually more appropriate for **ML** than **NF**, because they essentially involve the notion of being an element. Nevertheless, because all **NF**-“sets” are elements, the addition of this notion is inessential when applied to **NF**, and defining the abbreviations this way enables us to forego revising them in our later discussion of **ML**. The abbreviations, accordingly, are as follows:

$$x \in \hat{x}\varphi =_{df} \exists z (x \in z) \wedge \varphi, \quad (D1)$$

$$z = \hat{x}\varphi =_{df} \forall x (x \in z \leftrightarrow x \in \hat{x}\varphi), \quad (D2)$$

$$\hat{x}\varphi \in y =_{df} \exists z (z = \hat{x}\varphi \wedge z \in y). \quad (D3)$$

Now let $\bar{}$ be a one-to-one mapping of the individual and 1-place predicate variables onto the individual variables. We define a translation function s on the basis of this mapping by means of a recursion on the meaningful expressions of monadic λHST^* as follows:⁹

⁹Because we want to both use and mention \in here, we will take it as an autonymous sign, i.e., as one that serves to name itself.

1. if $a, b \in \mathbf{ME}_0$, then $s(a = b) = (s(\bar{a}) = s(\bar{b}))$,
2. if $a \in \mathbf{ME}_0$, and $\pi \in \mathbf{ME}_2$, then $s(\pi(a)) = (s(\bar{a}) \in s(\bar{\pi}))$,
3. $s([\lambda x \varphi]) = \hat{y}s(\varphi)$, where $y = \bar{x}$,
4. $s(\neg \varphi) = \neg s(\varphi)$,
5. $s(\varphi \rightarrow \psi) = (s(\varphi) \rightarrow s(\psi))$,
6. $s(\forall u \varphi) = \forall \bar{u}s(\varphi)$,
7. if $\varphi \in \mathbf{ME}_1$, then $s([\lambda \varphi]) = \hat{x}(s(\varphi) \wedge x = x)$.

We observe that if φ is a monadic wff of λHST^* that is an instance of either (A0*), (A1*), (A2*), (A3*), (LL*) or (Q^*), then by definition $s(\varphi)$ is provable in standard first-order predicate logic with identity, and therefore $s(\varphi)$ is a theorem of **NF**. On the other hand, if φ is an instance of ($HSCP_\lambda^*$), then $s(\varphi)$ is of the form $\exists z(\hat{x}\psi = z)$, and therefore by (D2), (D1) and the stratified comprehension principle of **NF** together with the fact that every **NF**-“set” is an element, it follows that $s(\varphi)$ is a theorem of **NF**. Similarly, that the s -transform of every instance of ($\lambda\text{-Conv}^*$) is a theorem of **NF** also follows from (D1) and the fact that every **NF**-“set” is an element. If φ is an instance of (Id_λ^*), then $s(\varphi)$ is of the form $\hat{x}(x \in a) = a$, and therefore by (D2) and (D1), it follows that $s(\varphi)$ is a theorem of **NF**. A similar observation applies to the s -transforms of instances of (Ext^*), moreover, and therefore since (MP) and (UG) preserve theoremhood in **NF**, it follows that the s -transform of each theorem of monadic $\lambda HST^* + (Ext^*) + (Q^*)$ is a theorem of **NF**. Putting our two results together, we can now state the following metatheorem:

Metatheorem 1. **NF** is equiconsistent with monadic $\lambda HST^* + (Ext^*) + (Q^*)$; and therefore because monadic $\lambda HST^* + (Ext^*) + (Q^*)$ is equiconsistent with the full system, **NF** is equiconsistent with $\lambda HST^* + (Ext^*) + (Q^*)$.

Before concluding this Section, it is worth noting that Ernst Specker has shown that **NF** is equiconsistent with the theory of simple types as a theory of classes if we add to the latter the assumption that all of the classes of any one type can be correlated one-to-one with the classes of the next succeeding type and that all of the classes of urelements can be correlated one-to-one with these urelements—an assumption that Specker calls “complete typical ambiguity” (cf. [23], p. 118). Such an assumption, needless to say, runs directly counter to the idea of sets as classes that are composed of their members, and therefore it fails to explain in what sense **NF** is to be viewed as a theory of sets. It does not run counter to the idea of **NF**-“sets” as concept-correlates, on the other hand, and in fact, given the assumption that every object is a concept-correlate, it conforms perfectly to our generalized form of Frege’s double-correlation thesis for unsaturated higher-level (unary) concepts. In this regard, we maintain, it is more appropriate to identify **NF**-“sets” with the concept-correlates of monadic

$\lambda HST^* + (Ext^*) + (Q^*)$ as a reconstruction of Frege's double-correlation thesis than to construe **NF** as a first-order reconstruction of the theory of types as a theory of classes.

6 On Taking Urelements Seriously

The idea that every object is a concept-correlate, i.e., Quine's thesis, is really not very plausible, and of course we do not recommend its adoption here. But because (Q^*) is equivalent to (Ext) in our reconstructed Fregean system, then dropping (Q^*) as an assumption means dropping (Ext) as well. Note, however, that even without (Q^*) the following restricted form of (Ext) is provable in $\lambda HST^* + (Ext^*)$:

$$\forall x \forall y (\exists z (z \in x) \wedge \forall z [z \in x \leftrightarrow z \in y] \rightarrow x = y). \quad (Ext')$$

Now the result of modifying **NF** by replacing (Ext) by (Ext') is precisely the system **NFU** (New Foundations with Urelements) described by Jensen in [14]. In that regard, accordingly, what the above observation shows is that **NFU** is contained in $\lambda HST^* + (Ext^*)$ in the same sense in which **NF** is contained in $\lambda HST^* + (Ext^*) + (Q^*)$. The question then arises as to whether monadic $\lambda HST^* + (Ext^*)$ can also be shown to be consistent relative to **NFU**.

The answer is affirmative, but in our proof we will proceed through an intermediary system **NFU'**, which is obtained from **NFU** by supplementing it with an individual constant Λ for the empty class. This means adding the axiom:

$$\neg \exists x (x \in \Lambda),$$

and modifying the stratified comprehension principle of **NFU** as follows:

$$\exists y ([\exists x (x \in y) \vee y = \Lambda] \wedge \forall x [x \in y \leftrightarrow \varphi]), \quad (CP-NFU')$$

where φ is a stratified first-order wff of **NFU'** in which y does not occur free. This modification of **NFU** is needed because in $\lambda HST^* + (Ext^*)$ there is a unique empty class (as a concept-correlate), whereas in **NFU**, because of the switch from (Ext) to (Ext') , there could be more than one empty class. Of course, because $(CP-NFU')$ implies the original comprehension principle of **NF** and **NFU**, then it follows that **NFU** is contained in **NFU'**. To prove that **NFU'** can be translated into **NFU**, on the other hand, we consider the following function defined recursively on the first-order wffs of **NFU'**:

1. $t(x \in y) = (x \in y)$,
2. $t(x \in \Lambda) = (x \neq x)$,
3. $t(\Lambda \in x) = \exists y [\neg \exists z (z \in y) \wedge y \in x]$,
4. $t(x = y) = (x = y)$,

5. $t(x = \Lambda) = t(\Lambda = x) = \neg \exists z(z \in x)$,
6. $t(\neg \varphi) = \neg t(\varphi)$,
7. $t(\varphi \rightarrow \psi) = [t(\varphi) \rightarrow t(\psi)]$,
8. $t(\forall x \varphi) = \forall x t(\varphi)$.

Note that the t -transform of any instance of $(CP\text{-}NFU')$ is a wff of **NFU** of the form:

$$\exists y([\exists x(x \in y) \vee \neg \exists x(x \in y)] \wedge \forall x[x \in y \leftrightarrow \varphi]),$$

where φ is a stratified wff in which y does not occur free, and therefore t translates instances of $(CP\text{-}NFU')$ into wffs that are trivially equivalent to the comprehension principle of **NFU**. But then because $t(Ext) = (Ext)$ and $t(\neg \exists x(x \in \Lambda)) = \neg \exists x(x \neq x)$, it follows that every theorem of **NFU'** is translated by t into a theorem of **NFU**. In other words, **NFU'** is equiconsistent with **NFU**.

To show that monadic $\lambda HST^* + (Ext^*)$ is consistent relative to **NFU'**, we return to the translation function s defined in Section 5, except that clause (6) in the definition of s is revised as follows:

- (6a) $s(\forall x \varphi) = \forall \bar{x} s(\varphi)$,
- (6b) $s(\forall F \varphi) = \forall \bar{F}[\exists z(z \in \bar{F}) \vee \bar{F} = \Lambda \rightarrow s(\varphi)]$.

Needless to say, the proof that s translates every theorem of monadic $\lambda HST^* + (Ext^*)$ into a theorem of **NFU'** is the same as our earlier proof regarding the translation of monadic $\lambda HST^* + (Ext^*) + (Q^*)$ into **NF**. It follows, of course, that the product of the translations t and s (i.e., $t \circ s$) translates every theorem of monadic $\lambda HST^* + (Ext^*)$ into a theorem of **NFU**. Accordingly, because for reasons already indicated we are able to drop the restriction to *monadic* $\lambda HST^* + (Ext^*)$, we have now proved the following metatheorem:

Metatheorem 2. $\lambda HST^* + (Ext^*)$ is equiconsistent with **NFU**.

Now aside from the above result, what makes **NFU** particularly interesting here are certain results of Jensen regarding both the consistency of **NFU** and the status of the axioms of infinity and choice in **NFU**. In particular, Jensen has shown that the axiom of infinity (Inf) is not provable in **NFU**, and that therefore the axiom of choice (AC) is not refutable therein. (Specker's proof that (AC) is disprovable in **NF** makes heavy use of (Ext)) This means that **NFU** is essentially weaker than **NF**, and therefore that (Q^*) in particular is not provable in $\lambda HST^* + (Ext^*)$. In addition, by the translation functions already defined, it follows that **NFU** + (Inf) + (AC) is equiconsistent with $\lambda HST^* + (Ext^*) + (Inf^*) + (AC^*)$ (cf. [4], Sections 5 and 8).

Another important result that Jensen has established is that **NFU** + (Inf) , + (Inf) + (AC) is consistent if weak Zermelo set theory is consistent (+ (Inf) , + (Inf) + (AC)). Weak Zermelo set theory (or what we will call weak **Z**) is

the restriction of Zermelo set theory (without (Inf)) to those instances of the *Aussonderungsaxiom* in which all quantifiers in the comprehension clause are limited or restricted, i.e., in which all quantifiers have the form $(\forall x \in y)\varphi$ or $(\exists x \in y)\varphi$. It should be noted, however, that Jensen's proof is semantical; that is, it does not transform a proof in **NFU** into a proof in weak **Z** but rather constructs a model of **NFU** in terms of a model of weak **Z**. Nevertheless, weak **Z** is particularly interesting here because Jensen has also shown (again by a semantical proof) that weak **Z** $(+(Inf), +(Inf) + (AC))$ is equiconsistent with the theory of simple types (**ST**) as a theory of classes (plus an axiom of infinity or axioms of infinity and choice in type-theoretical terms).

In [4], however, by using a Fregean semantics for $\lambda HST^* + (Ext^*)$ (in which general models are supplemented with a function correlating objects in the domain of discourse with values of the predicate variables), it was shown that a Fregean model of $\lambda HST^* + (Ext^*)$ $(+Inf^*, +(Inf^*) + (AC^*))$ can be reconstructed into a general model of **ST** (with corresponding type-theoretical axioms of infinity and choice). Accordingly, where **A** and **B** are arbitrary formal systems and $(\mathbf{A} \Rightarrow \mathbf{B})$ means that a (general) model of **B** can be constructed in terms of a (general) model of **A**, and $(\mathbf{A} \rightarrow \mathbf{B})$ means that **B** can be translated into **A** so that a proof in **B** can be transformed into a proof in **A**, we can represent the situation described above as follows (noting that the connections hold with (Inf) or $(Inf) + (AC)$ and their counterparts as well).

$$\begin{array}{ccc} \text{weak } \mathbf{Z} & \Longleftrightarrow & \mathbf{ST} \\ \Downarrow & & \Downarrow \\ \mathbf{NFU} & \leftrightarrow & \lambda HST^* + (Ext^*) \end{array}$$

In other words, all of these formal systems are equiconsistent with one another.

Metatheorem 3. Weak **Z**, **ST**, **NFU** and $\lambda HST^* + (Ext^*)$, with or without axioms of infinity and choice, are all equiconsistent with each other.

it should perhaps be emphasized before concluding this Section, incidentally, that the equiconsistency between all of these systems does not show that the restricted form of the iterative concept of set (as represented by **ST** and weak **Z**) coincides with the logical concept of class (as represented by **NFU** and $\lambda HST^* + (Ext^*)$). Rather, the fact that some of the consistency proofs in question are semantical (and may involve general models at that) indicates that equiconsistency only shows that models of the one concept (as represented by one of the systems in question) can be constructed in terms of models of the other.

7 An Alternative Modification of Frege's Double-Correlation Thesis

Now there is another way of modifying Frege's double-correlation thesis so that Russell's paradox does not follow. In particular, instead of modifying the thesis as it applies directly to the positing of first-level concepts and only indirectly to the positing of concept-correlates— i.e., instead of modifying the thesis as represented by the original comprehension principle (CP_λ^*) — we might instead modify only its application to the positing of concept-correlates. The framework for such a modification, incidentally, is implicit, at least in part, in Frege's discussion of Russell's paradox when he suggests that we might “suppose there are cases where an unexceptional concept has no class answering to it as its extension” ([17], p. 128). That is, instead of assuming that each singular term actually denotes (a value of the bound individual variables) — i.e., instead of assuming axiom (A3*) — we might allow that some singular terms, and certain nominalized predicates in particular, are denotationless. The modification in question, in other words, is to be formulated in a logic that is free of existential presuppositions regarding singular terms.

Let us return, accordingly, to the second-order logic with nominalized predicates formulated in Section 3, i.e., the logic that we took as a counterpart to Frege's own original system. λ -abstracts need not be homogeneously stratified in this logic, it will be remembered, and although this returns us once again to Russell's paradox, a contradiction is avoided by adopting the suggestion in question, i.e., by allowing for denotationless singular terms. We may do so, moreover, first by replacing (A3*) by the following weaker, but clearly unexceptional, “law of logic”:

$$(A3^{**}) \quad \forall x \exists y (x = y).$$

Secondly, because without (A3*) the identity law ($a = a$), where a is an arbitrary singular term, is no longer derivable, we will have to add ($a = a$) as an axiom schema as well. Finally, we will need to replace ($\lambda\text{-Conv}^*$) by the presupposition-free form of λ -conversion:

$$[\lambda x_1 \dots x_n \varphi](a_1, \dots, a_n) \leftrightarrow \exists x_1 \dots \exists x_n (a_1 = x_1 \wedge \dots \wedge a_n = x_n \wedge \varphi) \quad (\exists/\lambda\text{-Conv}^*)$$

where no x_i is free in any a_j , for all i, j such that $1 \leq i, j \leq n$.

Now with the replacement of (A3*) by (A3**) and ($a = a$) and of ($\lambda\text{-Conv}^*$) by ($\exists/\lambda\text{-Conv}^*$), the principle (UI_1^*) of universal instantiation of a singular term for an individual variable is no longer derivable except in the following qualified form:

$$\exists y (a = y) \rightarrow [\forall x \varphi \rightarrow \varphi(a/x)], \quad (\exists/UI_1^*)$$

where a is any singular term that is free for x in φ and in which y has no free occurrences. Note that with this qualification what follows from Russell's paradox is not a contradiction but only that the predicate $[\lambda x \exists G(x = G \wedge \neg G(x))]$ is denotationless in its occurrences as a singular term; i.e., instead of a

contradiction, only

$$\neg \exists y(y = [\lambda x \exists G(x = G \wedge \neg G(x))])$$

is provable. In other words, even though the unqualified comprehension principle (CP_λ^*) posits the existence of an unsaturated concept corresponding to $[\lambda x \exists G(x = G \wedge \neg G(x))]$ as a predicate, nevertheless by Russell's argument it is provable in the system in question that there can be no saturated object corresponding to $[\lambda x \exists G(x = G \wedge \neg G(x))]$ as a singular term; or in Frege's words here is a case "where an unexceptional concept has no class answering to it as its extension" (*ibid.*).

That some nominalized predicates must be denotationless in the present system, incidentally, does not show that all must be denotationless (cf. [5], Section 14 for a discussion of the latter option). Indeed, if arithmetic is to be derivable in the framework in question, then we must assume that at least some concepts have concept-correlates. Of course, such an assumption can easily be given in terms of the existence conditions for sets in one or another version of set theory; but the result, while perhaps interesting as a second-order set theory, will mean giving up the logical concept of class. That is, since the existence conditions in question are in accordance with the iterative concept of set, and therefore with the limitation of size doctrine, the resulting framework would really only be concerned with classes as composed of their members and not with classes in the logical sense. Note, incidentally, that it would be philosophically incoherent to drop the axiom of extensionality in such a second-order set theory (or even in a first-order set theory for that matter) and call the result *a theory of properties* (as abstract individuals). For even without an axiom of extensionality, the abstract individuals characterized by set-theoretical axioms are still posited in accordance with the iterative concept of set, and in that sense they are still composed of their members. Thus, if the objects denoted by nominalized predicates are to be "logical objects", which in the presence of an axiom of extensionality means that they are to be classes in the logical sense, then the relevant assumption regarding the positing of such objects should be in accordance with Frege's double-correlation thesis and not with the iterative concept of set.

Now in regard to Frege's double-correlation thesis, let us note that because we are retaining the full unrestricted comprehension principle (CP_λ^*), then we are in effect also retaining the unrestricted form of Frege's double-correlation thesis *insofar as it applies to the positing of first-level concepts*. That is, we are retaining that part of the thesis that posits first-level concepts corresponding to unequal-leveled or inhomogeneous higher-level relations no less so than to homogeneous or equal-leveled relations. Thus, corresponding to Frege's unequal second-level relation of subsumption or predication, there is in the system as so far described a first-level relation of subsumption posited by (CP_λ^*); i.e.,

$$\exists R^2(R = [\lambda xy \exists G(x = G \wedge G(y))])$$

is provable in the system in question. Of course, by Russell's argument this same predicate when nominalized must be denotationless; i.e., in its occurrences as a

singular term it must fail to denote (a value of the bound individual variables). Or, in other words, as applied to concept-correlates, Frege's double-correlation thesis must be restricted, and in particular it is not to apply to all of the first-level concepts corresponding to or otherwise Involving inhomogeneous higher-level relations. The assumption we shall make here is that it is to apply *at least* (but by no means therefore only) to all of the first-level concepts corresponding to higher-level unsaturated concepts (including relations) that are homogeneously stratified; i.e., *at least* to all or the concepts that have concept-correlates in λHST^* .

In order to formulate our assumption as an axiom schema, we will say that a meaningful expression (as defined in terms of the logical grammar of Section 3) is *bound to individuals* if and only if for all predicate variables F , and all wffs φ , if $\forall F\varphi$ is a wff occurring in ξ , then for some individual variable x and some wff ψ , φ is the wff $[\exists x(F = x) \rightarrow \psi]$. Thus, to be bound to individuals, every predicate quantifier occurring in ξ must refer only to those first-level concepts posited by (CP_λ^*) that have corresponding concept-correlates. Our assumption regarding the existence of concept-correlates may now be formulated as the following axiom schema:

$$\exists y(a_1 = y) \wedge \dots \wedge \exists y(a_k = y) \rightarrow \exists y(y = [\lambda x_1 \dots x_n \varphi]), \quad (\exists/HSCP_\lambda^*)$$

where $[\lambda x_1 \dots x_n \varphi]$ is homogeneously stratified, φ is bound to individuals, y is an individual variable not occurring in φ , and a_1, \dots, a_k are all of the variables (or nonlogical constants) occurring free in $[\lambda x_1 \dots x_n \varphi]$.¹⁰

The axiom schemas of our present system, accordingly, are $(A0^*)$, $(A1^*)$, $(A2^*)$, $(A3^*)$, $(a = a)$, where a is any singular term, (LL^*) , (CP_λ^*) , $(\exists/\lambda-Conv^*)$, (ld_λ^*) and now $(\exists/HSCP_\lambda^*)$ as well.¹¹ (We retain modus ponens and universal generalization as our only inference rules.) Because of its relation to our earlier system λHST^* , we will refer to this system here-after as HST_λ^* .

Now it is clear that HST_λ^* is not a conservative extension of λHST^* , because whereas, by $(A3^*)$, $\forall F \exists x(F = x)$ is provable in λHST^* , this same wff, as witness the predicate $[\lambda x \exists G(x = G \wedge \neg G(x))]$, is actually disprovable in HST_λ^* . Nevertheless, because every wff of λHST^* is provably equivalent, again by $(A3^*)$, to a wff that is bound to individuals, HST_λ^* may be said to contain λHST^* in the sense of the following lemma (which is easily seen to hold).

Lemma: If φ is a wff of λHST^* that is bound to individuals, y is an individual variable not occurring in φ , and a_1, \dots, a_k are all of the variables (or nonlogical constants) occurring free in φ , then $\vdash_{\lambda HST^*} \varphi$ only if $\vdash_{HST_\lambda^*} \exists y(a_1 = y) \wedge \dots \wedge \exists y(a_k = y) \rightarrow \varphi$.

Now because an object that is denoted by a λ -abstract of $\lambda HST^*(+(Ext^*), +(Ext^*) + (Q^*), \text{ etc.})$ is, by $(A3^*)$, also denoted by a λ -abstract that is bound

¹⁰We understand the "conditional" posited in $(\exists/HSCP_\lambda^*)$ to be just the consequent if $k = 0$, i.e., if the λ -abstract in question contains no free variables (or nonlogical constants).

¹¹The rewrite law cited in footnote 5 is needed here as well if neither (Ext^*) nor $(\Box Ext^*)$ is assumed as a supplementary axiom schema.

to individuals, it follows, by $(\exists/HSCP_\lambda^*)$, that every object that is a concept-correlate in $\lambda HST^*((+Ext^*), +(Ext^*)+(Q^*), \text{etc.})$ is also a concept-correlate in $HST_\lambda^*(+(Ext^*), +(Ext^*)+(Q^*), \text{etc.})$; and therefore because **NFU** and **NF** are contained, respectively, in $\lambda HST^*+(Ext^*)$ and $\lambda HST^*+(Ext^*)+(Q^*)$, then **NFU** and **NF** are also contained (in the sense of the above lemma) in $\lambda HST^*+(Ext^*)+(Q^*)$ and $HST_\lambda^*+(Ext^*)+(Q^*)$, respectively. Moreover, restricting ourselves to pure wffs (i.e., wffs in which no predicate or individual constants occur), it follows by the above lemma that every sentence (wff with no free variables) of λHST^* that is bound to individuals and provable in $\lambda HST^*(+(Ext^*), +(Ext^*)+(Q^*), \text{etc.})$ is also provable in $HST_\lambda^*(+(Ext^*), +(Ext^*)+(Q^*), \text{etc.})$ as well; i.e., that proofs in λHST^* and its extensions can be transformed into related proofs in HST_λ^* and its corresponding extensions. It follows, accordingly, that if $HST_\lambda^*(+(Ext^*), +(Ext^*)+(Q^*), \text{etc.})$ is consistent, then so is $\lambda HST^*(+(Ext^*), +(Ext^*)+(Q^*), \text{etc.})$.

It is also true, moreover, that if $\lambda HST^*+(Ext^*)$ or any of its extensions is consistent, then so is $HST_\lambda^*+(Ext^*)$ and its corresponding extensions. The proof of this claim, however, is semantical and does not provide a method whereby a proof in $HST_\lambda^*+(Ext^*)$ of a wff of λHST^* can be transformed into a proof in $\lambda HST^*+(Ext^*)$. (Cf. [10], p.169, footnote 1, for the related claim regarding the impossibility of transforming proofs in **ML** into proofs in **NF**.) That is, the proof depends on developing a Fregean semantics for these systems that we will not go into here (but cf. [6], Chapter 6, Section 4, for the details of such a semantic proof). Nevertheless, on the basis of the details of that semantical proof it can be shown that a sentence of λHST^* that is bound to individuals is provable in $\lambda HST^*+(Ext^*)$ if, as well as only if, it is provable in $HST_\lambda^*+(Ext^*)$; and therefore it follows that the one system (and any of its extensions) is consistent if and only if the other (and its corresponding extensions) is consistent. Putting this result together with those already noted in the preceding Sections, we have the following in metatheorem.

Metatheorem 4. $HST_\lambda^*+(Ext^*)$, with or without axioms of infinity and choice, is equiconsistent with $\lambda HST^*+(Ext^*)$, weak **Z**, **ST** and **NFU**, all with or without corresponding axioms of infinity and choice; and $HST_\lambda^*+(Ext^*)+(Q^*)$ is equiconsistent with **NF**.

8 Ultimate Classes and the Similarity of ML with $HST_\lambda^*+(Ext^*)+(Q^*)$

Having given a natural motivation for both **NF** and **NFU** in terms of our first modification of Frege's double-correlation thesis, let us now see how our second or alternative modification of Frege's thesis can be used to explain **ML** as well. In **ML**, it will be remembered, sets are not just classes that are composed of their members; rather, they are also classes that are elements. That is, in **ML**, which is also an applied first-order theory with \in as its only primitive predicate

constant, the notion of a set is defined as follows:

$$Set(x) =_{df} \exists z(x \in z)$$

Classes which are not sets are then referred to as *proper* or *ultimate classes*. (We will follow Quine [19] and call them ultimate classes hereafter.)

Of course, both the sets and ultimate classes of **ML** are assumed by Quine to be classes as composed of their members, which on Quine's view only means that the axiom of extensionality, (*Ext*), holds for both. In other words, according to Quine, the being of an ultimate class is no different from that of a class that is an element; they are both classes as composed of their members. But then why can't ultimate classes be elements? That a contradiction, namely Russell's paradox, would ensue other-wise surely cannot be taken as an explanation of why ultimate classes cannot be members of, say, finite classes (cf. [10], p. 142).

The answer, we maintain, is that ultimate classes are really not objects to begin with, but are rather unsaturated concepts that have no classes answering to them as their extensions. That is, not being objects to begin with, we can hardly demand an explanation of why ultimate classes cannot be elements, for only an object can be a member, be it of a finite or an infinite class. The "sets" of **ML**, accordingly, are really not sets in the sense of classes that are composed of their members but are rather classes in the logical sense; i.e., they are classes that have their being as concept-correlates, and in particular they are the concept-correlates of monadic $HST_\lambda^* + (Ext^*) + (Q^*)$.

To see that this is so, let us note that by $(\exists/\lambda-Conv^*)$,

$$\vdash_{HST_\lambda^*} [\lambda x F(x)](a) \rightarrow \exists x(a = x),$$

where x is not free in a ; and therefore by (Id_λ^*) and (UG),

$$\vdash_{HST_\lambda^*} \forall F[F(a) \rightarrow \exists x(a = x)].$$

In other words, according to HST_λ^* only real objects (in the sense of the values of the bound individual variables) can fall under concepts (which is not the same as to say that wffs containing denotationless singular terms cannot be true). Accordingly, assuming the same analysis of membership as given in Section 5, it follows that only real objects can be elements:

$$\vdash_{HST_\lambda^*} \exists x(a \in x) \rightarrow \exists x(a = x).$$

The converse also holds, moreover; for even though $(a = a)$ is true when a is denotationless, nevertheless by $(\exists/\lambda-Conv^*)$,

$$\vdash_{HST_\lambda^*} \exists x(a = x \wedge x = x) \rightarrow [\lambda x(x = x)](a),$$

and therefore because, by (CP_λ^*) , $[\lambda x(x = x)]$ stands for a concept, it follows by the definition of membership that

$$\vdash_{HST_\lambda^*} \exists x(a = x) \rightarrow \exists x(a \in x).$$

That is, according to HST_λ^* only elements can be real objects.

Now the ultimate classes of **ML** are really not objects, we have said, but are rather unsaturated concepts — or that in any case is how they should be viewed if we are to explain or provide a natural motivation for the classes of **ML**. Of course, in **ML** ultimate classes are represented as if they were real objects, because they are there taken as values of the bound individual variables. But that only means that the individual variables of **ML** should be reconstrued as one-place predicate variables so that all of the “classes” of **ML**, ultimate or otherwise, can be seen for what they are, namely as unsaturated concepts; and then, only if they are also “elements” in **ML** should they be viewed as having objects, namely concept-correlates, corresponding to them. It is only in this way, we maintain, that we can explain why ultimate “classes” cannot be elements, as well as provide a natural motivation for the classes of **ML** that are elements. In other words, where $q(\psi)$, for each wff ψ of **ML**, is the result of replacing each individual variable in ψ by a one-place predicate variable (where distinct individual variables are replaced by distinct predicate variables), our claim is that $q(\psi)$ is a theorem of monadic $HST_\lambda^* + (Ext^*) + (Q^*)$ whenever ψ is a theorem of **ML**.

Note in this regard that where ψ is an instance of the comprehension principle *202 of **ML** for classes, ultimate or otherwise, then $q(\psi)$ is of the following form:

$$\exists F \forall G (G \in F \leftrightarrow \exists H (G \in H) \wedge \varphi).$$

In HST_λ^* , however, not only does (CP^*) follow from (CP_λ^*) , but, given our analysis of membership and the above observation that all and only real objects are elements, we also have the following:

$$\vdash_{HST_\lambda^*} \exists F \forall x (x \in F \leftrightarrow \exists z (x \in z) \wedge \varphi).$$

Therefore, by (Q^*) , the q -transform of each instance of the comprehension principle *202 of **ML** is a theorem of monadic $HST_\lambda^* + (Q^*)$.

Now aside from (Ext) , the q -transform of which is easily seen to be derivable in HST_λ^* on the basis of (Ext^*) and (Q^*) , the remaining axiom schema of **ML** is *200 which depends on the notion of normalcy.¹² In particular, a wff ψ of **ML** is said to be *normal* if all of the bound variables in ψ are restricted to elements; i.e., if in all of the subwffs of ψ of the form $\forall x \varphi$, φ is of the form $[\exists y (x \in y) \rightarrow \chi]$. Axiom schema *200 can then be stated as follows (cf. [18], p. 162):

***200:** If ψ is a stratified, normal wff of **ML**, y is new to ψ , and x, z_1, \dots, z_n are all of the distinct individual variables occurring free in ψ , then:

$$\vdash_{\mathbf{ML}} \exists y (z_1 \in y) \wedge \dots \wedge \exists y (z_n \in y) \rightarrow \exists y (\exists z (y \in z) \wedge \forall x [x \in y \longleftrightarrow \exists z (x \in z) \wedge \psi]).$$

¹²This notion of normalcy comes from Gödel [12], p. 12. Its application to **ML** was first given in Wang [24].

Needless to say, but by our above observation that all and only real objects are elements in HST_λ^* , it is clear that a wff ψ of **ML** is normal if and only if $q(\psi)$ is provably equivalent in HST_λ^* to a wff that is bound to individuals. Applying the same observation to $(\exists/HSCP_\lambda^*)$, moreover, we have the following result:

$$\vdash_{HST_\lambda^*} \exists y(a_1 \in y) \wedge \dots \wedge \exists y(a_n \in y) \rightarrow \exists y(y = [\lambda x\varphi]),$$

where y is new to φ , and x, a_1, \dots, a_n are all the distinct variables occurring free in φ , φ is bound to individuals and $[\lambda x\varphi]$ is homogeneously stratified. But, again by the same observation, and by the same sort of argument that shows that (CP^*) is a consequence of (CP_λ^*) , the following is easily seen to hold:

$$\vdash_{HST_\lambda^*} \exists y(y = [\lambda x\varphi]) \rightarrow \exists y(\exists z(y \in z) \wedge \forall x[x \in y \leftrightarrow \exists z(x \in z) \wedge \varphi]).$$

Accordingly, by putting these results together and then applying (Q^*) , it follows that the q -transform of every instance of *200 is a theorem of monadic $HST_\lambda^* + (Ext^*) + (Q^*)$; and this completes our proof that the q -transform of a theorem of **ML** is a theorem of monadic $HST_\lambda^* + (Ext^*) + (Q^*)$.

We can also show, as might well be expected by now, that every theorem of monadic $HST_\lambda^* + (Ext^*) + (Q^*)$ can be translated into a theorem of **ML**. For this purpose, we need only return to the translation function s defined in Section 5, except that clause (6) is revised as follows:

$$\begin{aligned} (6a) \quad & s(\forall x\varphi) = \forall \bar{x}[\exists z(x \in z) \rightarrow s(\varphi)], \\ (6b) \quad & s(\forall F\varphi) = \forall \bar{F}s(\varphi). \end{aligned}$$

The proof that s translates each axiom of monadic $HST_\lambda^* + (Ext^*) + (Q^*)$ into a theorem of **ML** is similar to the proof that s as originally defined in Section 5 translates the axioms of monadic $\lambda HST_\lambda^* + (Ext^*) + (Q^*)$ into theorems of **NF**, except that now the element conditions that are part of the abbreviations (D1)–(D3) are essential to the proof. Putting these two results together, we can now state the following metatheorem (keeping in mind that monadic HST_λ^* is equiconsistent with the full system).

Metatheorem 5.

- (a) If ψ is a theorem of **ML**, then $q(\psi)$ is a theorem of $HST_\lambda^* + (Ext^*) + (Q^*)$;
- (b) if φ is a theorem of monadic $HST_\lambda^* + (Ext^*) + (Q^*)$, then $s(\varphi)$ is a theorem of **ML**; and
- (c) therefore $HST_\lambda^* + (Ext^*) + (Q^*)$ is equiconsistent with **ML**.

Finally, let **MLU'** (**ML** with urelements) be the result of replacing (Ext) in **ML** by (Ext') , adding Λ as an individual constant together with the axiom,

$$\neg \exists x(x \in \Lambda),$$

and modifying the comprehension principle *202 of **ML** as follows:

$$\exists y([\exists x(x \in y) \vee y = \Lambda] \wedge \forall x[x \in y \leftrightarrow \exists z(x \in z) \wedge \varphi]). \quad (CP\text{-}MLU')$$

Then, by revising clause (6b) above of the translation functions as follows,

$$(6b) \quad s(\forall F \varphi) = \forall \bar{F} [\exists z (z \in \bar{F}) \vee \bar{F} = \Lambda \rightarrow s(\varphi)],$$

we can repeat the above argument and show that s translates every theorem of $HST_\lambda^* + (Ext^*) + (Q^*)$ into a theorem of \mathbf{MLU}' .

What cannot be shown, however, is that the theorems of \mathbf{MLU}' can be translated into theorems of $HST_\lambda^* + (Ext^*)$. E.g., the q -transform of (Ext') is not a theorem of $HST_\lambda^* + (Ext^*)$, because without (Q^*) , the conditional

$$\forall H (H \in F \leftrightarrow H \in G) \rightarrow \forall x (x \in F \leftrightarrow x \in G)$$

is not provable in $HST_\lambda^* + (Ext^*)$; i.e., without (Q^*) neither (Ext) nor the q -transform of (Ext') is provable in HST_λ^* on the basis of (Ext^*) . On the other hand, if we drop the idea of correlating the individual variables of \mathbf{ML} with the predicate variables of monadic $HST_\lambda^* + (Ext^*)$, then, however defined, the transform of $(\mathbf{CP-MLU}')$ will not be provable in $HST_\lambda^* + (Ext^*)$, because it will in that case posit the existence of a concept-correlate corresponding to any given concept, which of course is impossible if $HST_\lambda^* + (Ext^*)$ is consistent. In addition, such an alternative would mean giving up the whole idea of explaining why ultimate classes cannot be elements. The fact, in other words, that \mathbf{MLU}' cannot be translated into $HST_\lambda^* + (Ext^*)$ indicates that \mathbf{MLU}' lacks a natural motivation, especially because, as already noted in Section 7, every class in \mathbf{NFU} turns out to be a concept-correlate in $HST_\lambda^* + (Ext^*)$. Accordingly, if we are to drop Quine's thesis, whether in the form (Ext) or (Q^*) , then the appropriate framework is not \mathbf{MLU}' but $HST_\lambda^* + (Ext^*)$ instead.

9 On Mathematical Induction and the Class of Fregean Natural Numbers

Quine's principal reason for proposing \mathbf{ML} , it should be noted, is that, unlike the situation in \mathbf{NF} , mathematical induction can be proved in \mathbf{ML} without restriction to stratified wffs (cf. [18], p. 165; [19], p. 300). This is because in both \mathbf{NF} and \mathbf{ML} it is Frege's definition of the natural numbers that is both natural and most appropriate (cf. [19], p. 289); and, of course, that is as it should be if both \mathbf{NF} and \mathbf{ML} are to be explained in terms of Frege's double-correlation thesis. On this definition, at least as reconstructed in both \mathbf{NF} and \mathbf{ML} , a natural number is any class that belongs to *every class, ultimate or otherwise*, to which 0 belongs and that is closed under the successor relation. By definition, accordingly, the only form of mathematical induction that can be proved in \mathbf{NF} is one that is restricted to stratified wffs (even though not all of the wffs of \mathbf{NF} are stratified), because classes in general are specified in \mathbf{NF} only in terms of stratified wffs. In \mathbf{ML} , however, stratification is not necessary for the specification of classes in general, but only for the specification of classes that are elements. Thus, given Frege's definition of the natural numbers, it is only in \mathbf{ML} that the full form of mathematical induction is provable.

The problem, however, is that although the class of natural numbers is specifiable in terms of a stratified wff, and therefore can be proved to exist within **NF**, the same specification when given in **ML** proves the “existence” of the class of natural numbers only in the sense of its being a value of the bound individual variables, and not also in the more critical sense (at least for the purposes of mathematics) of its being an element. Or, given our reconstrual of the individual variables of **ML** as the predicate variables on monadic $HST_\lambda^* + (Ext^*) + (Q^*)$, it is only the *concept* of a natural number whose existence (as a value of the bound predicate variables) is assured, and not also the class of natural numbers as an object that is correlated with this concept.

Of course, the class of natural numbers whose existence is proved in **NF** can be proved to be an element in **ML**; but, *as specified in ML* (i.e., in terms of a wff that is both *stratified and normal*), this class does not validate the full form of mathematical induction. Nevertheless, because the same specification without the normalcy condition is even more restrictive (i.e., the wff without the normalcy condition implies the wff with the normalcy condition), then the class as specified with the normalcy condition is really the class of natural numbers only if it is identical with the class that is specified without the normalcy condition; and therefore, because the one is an element in **ML**, we may assume that the “other” is as well (cf. [19], p. 303f). In other words, according to Quine, we may assume as an additional axiom of **ML** the identity of the class of natural numbers as specified with the normalcy condition and as specified without this condition. In that way, one can prove in **ML** both the unrestricted form of mathematical induction and the elementhood of the class of natural numbers (as specified in terms of ultimate classes, i.e., without the normalcy condition).

There is an objection to adding such an axiom to **ML**, however, namely that insofar as “**ML** was proposed as an improvement of **NF** because **ML** has full-fledged mathematical induction whereas **NF** has induction for stratified formulae only, one may prefer to stick to **NF** since, once one starts adjoining to **ML** axioms as mentioned, one might as well add an axiom of full mathematical induction to **NF**” ([10], p. 171). Now it is important to note that such an objection, while it might apply to Quine’s own reason for proposing **ML** (especially when one attempts to view both **NF** and **ML** merely as theories of classes as composed of their members), completely misses its mark when **NF** and **ML** are explained in terms of Frege’s double-correlation thesis; i.e., when **NF** is viewed as contained in $\lambda HST + (Ext^*) + (Q^*)$ and **ML** is viewed (under the q -transformation) as contained in $HST_\lambda^* + (Ext^*) + (Q^*)$. Of course, in that case, we might just as well drop (Q^*) , or equivalently, (Ext) , as a form of Quine’s thesis and replace **NF** and **ML** by their preferred Fregean counterparts, $\lambda HST + (Ext^*)$ and $HST_\lambda^* + (Ext^*)$.

In reconsidering the problem from this perspective, in other words, we must first understand that the reason for proposing $HST_\lambda^* + (Ext^*)$ as an alternative to $\lambda HST + (Ext^*)$ is not because it has the full form of mathematical induction as a consequence of Frege’s definition of the natural numbers but rather because it is an alternative to $\lambda HST + (Ext^*)$ as a reconstruction of Frege’s double-

correlation thesis. Such an alternative, moreover, does not extend $\lambda HST + (Ext^*)$ by the introduction of a new notion, such as that of an ultimate class as opposed to the notion of a class as an element, because in both $\lambda HST + (Ext^*)$ and $HST_\lambda^* + (Ext^*)$ all classes are elements. Indeed, because the distinction between an ultimate class and a class that is an element is really that between a concept and a class, then, on our explanation no new notion has really been introduced in **ML** after all; for the distinction between concepts and the classes that are their extensions is as fundamental in $\lambda HST + (Ext^*)$ as it is in $HST_\lambda^* + (Ext^*)$.

Secondly, if $HST_\lambda^* + (Ext^*)$ is to be preferred to $\lambda HST + (Ext^*)$, then it is not so much because the unrestricted form of mathematical induction can be proved therein as because it leaves unmodified the full force of Frege's double-correlation thesis insofar as the latter applies to the positing of first-level concepts. In other words, whereas according to $\lambda HST + (Ext^*)$ there are no higher-level inhomogeneous relations, or at least none that can be correlated with the first-level concepts posited in $(HSCP_\lambda^*)$, the situation is completely otherwise in $HST_\lambda^* + (Ext^*)$, and in that regard there may well be more classes implicit in the latter framework than are acknowledged in the former, and therefore more than are posited in $(\exists/HSCP_\lambda^*)$. Thus, while Russell's paradox shows us that not all of the first-level concepts that correspond to or otherwise involve inhomogeneous higher-level relations have a corresponding concept-correlate, nothing follows to the effect that none of those concepts can have such a corresponding concept-correlate.

In particular, where S abbreviates the λ -abstract that stands for Frege's successor relation (cf. [19], p. 279), and

$$0 =_{df} [\lambda x \exists F(x = F \wedge \neg \exists y F(y))],$$

$$N =_{df} [\lambda x \forall F(F(0) \wedge \forall y \forall z[F(y) \wedge S(y, z) \rightarrow F(z)] \rightarrow F(x))].$$

then even though N is not bound to individuals (and therefore is not posited in $(\exists/HSCP_\lambda^*)$ as denoting a class), nevertheless, it seems plausible to assume

$$\exists x(x = N) \tag{N^*}$$

as an additional axiom. That is, it seems plausible to assume that as represented in $HST_\lambda^* + (Ext^*)$ the Fregean concept of a natural number has a class answering to it as its extension. It follows of course that in $HST_\lambda^* + (Ext^*) + (N^*)$ we have both the class of natural numbers as a concept-correlate and the unrestricted form of the principle of mathematical induction.

Note, incidentally, that where

$$N' =_{df} [\lambda x \forall w(0 \in w \wedge \forall y \forall z[y \in w \wedge S(y, z) \rightarrow z \in w] \rightarrow x \in w)],$$

then although we can prove the following identity wff in $HST_\lambda^* + (Ext^*)$,

$$N' = [\lambda x \forall F(\exists w(F = w) \wedge F(0) \wedge \forall y \forall z[F(y) \wedge S(y, z) \rightarrow F(z)] \rightarrow F(x))],$$

nevertheless we cannot prove $(N = N')$ in $HST_\lambda^* + (Ext^*)$ without assuming (N^*) . In other words, because $\forall x[N(x) \rightarrow N'(x)]$ is provable even in HST_λ^* , the class denoted by N' may contain objects other than the natural numbers as well; and if that is the case, then this alone would suffice as a reason to prefer $HST_\lambda^* + (Ext^*) + (N^*)$ over $\lambda HST^* + (Ext^*)$ even when the unrestricted principle of mathematical induction is added to the latter. For it is easily seen that $(N = N') \leftrightarrow (N^*)$ is provable in $HST_\lambda^* + (Ext^*)$.

Now we are not suggesting that $(N \neq N')$ is true in the framework represented by $HST_\lambda^* + (Ext^*)$, but only that one appropriate way to investigate this question is by considering additional axioms or principles regarding the positing of concept-correlates for first-level concepts that either correspond to or otherwise involve inhomogeneous higher-level relations, and which might thereby warrant assuming (N^*) , from which $(N = N')$ follows in $HST_\lambda^* + (Ext^*)$. The initial system HST_λ^* , in other words, is not proposed as an alternative to λHST^* because it allows for the unrestricted principle of mathematical induction, but rather because it captures the full force of Frege's double-correlation thesis at least as far as the positing of first-level concepts is concerned. The axiom $(\exists/HSCP_\lambda^*)$ then assures us that all of the concept-correlates (or classes, given (Ext^*)) that can be proved to exist in λHST^* are carried over into HST_λ^* , but by no means does it exhaust all of the possibilities as to which concepts can or cannot have concept-correlates.

In concluding, it should perhaps be noted that without Quine's thesis, whether in the form (Ext) or (Q^*) , the axiom of infinity is not provable in either $\lambda HST^* + (Ext^*)$ or $HST_\lambda^* + (Ext^*) + (N^*)$. Because such an axiom is needed at least for the purposes of logicism, it may be added to both of these systems in either one of the usual forms (cf. [4], Section 5) or as follows when relations are included as well:

$$\forall F^m \forall G^n (F \neq G), \quad (Inf^*)$$

where m, n are arbitrary natural numbers such that $m \neq n$.

In other words, if for all distinct natural numbers m and n the concept-correlate of any m -ary relation is other than the concept-correlate of any n -ary relation, then there are infinitely many (or at least a potential infinity) of concept-correlates. Such an assumption, it should be noted, is independent of the question of how many objects other than concept-correlates there are, and in particular of how many concrete objects there are.

References

- [1] G. Boolos, 'The iterative conception of set', *The Journal of Philosophy* 68 (1971), 215-231.
- [2] A. Church, 'Schröder's anticipation of the simple theory of types', *The Journal of Unified Science (Erkenntnis)* 9 (1939), 149-152.
- [3] N.B. Cocchiarella, 'A second-order logic of variable-binding operators', *Reports on Mathematical Logic* 5 (1975).3-18.

- [4] N.B. Cocchiarella, 'The theory of homogeneous simple types as a second-order logic', *Notre Dame Journal of Formal Logic* 20 (1979), 505–524.
- [5] N.B. Cocchiarella, 'Frege, Russell and Logicism: a logical reconstruction', *Frege Synthesized: Essays on the Philosophical and Foundational Work of Gottlob Frege*, Leila Haaparanta and Jaakko Hintikka, eds., D. Reidel Pub. Co., Dordrecht, 1986: 197–252.
- [6] N.B. Cocchiarella, *Logical Investigations of Predication Theory and the Problem of Universals*, vol. 2 of *Indices*, Bibliopolis Press, Naples, 1986.
- [7] G. Frege, *The Basic Laws of Arithmetic*, translated by M. Furth, University California Press, Berkeley, 1964.
- [8] G. Frege, *Posthumous Writings*, eds. H. Hermes, F. Kambartel and P. Kaulbach; translated by P. Long and R. White, Blackwell, Oxford, 1979.
- [9] G. Frege, *Translations from the Philosophical Writings of Gottlob Frege*, eds. P. Geach and M. Black, Blackwell, Oxford, 1952.
- [10] A. Fränkel, Y. Bar-Hillel, and A. Levy, *Foundations of Set Theory*, North-Holland, Amsterdam, 1973.
- [11] D. Gallin, *Intensional and Higher-Order Modal Logic*, North-Holland, Amsterdam, 1975.
- [12] K. Gödel, *The Consistency of the Continuum Hypothesis*, Princeton University Press, Princeton, 1940.
- [13] L. Henkin, 'Completeness in the theory of types', *Journal of Symbolic Logic* 15 (1950), 81–91.
- [14] R. Jensen, 'On the consistency of a slight(?) modification of Quine's New Foundations', *Synthese* 19 (1968), 250–263.
- [15] G. Kreisel, 'Two notes on the foundations of set-theory', *Dialectica* 23 (1969), 93–114.
- [16] R.M. Montague, *Formal Philosophy*, ed., R.H. Thomason, Yale University Press, New Haven, 1974.
- [17] W.V.O. Quine, 'New foundations for mathematical logic', *American Math. Monthly* 44 (1937), 70–80; reprinted with additions in *From a Logical Point of View*, Harvard Univ. Press, 2nd ed., Cambridge, 1961.
- [18] W.V.O. Quine, *Mathematical Logic*, Harvard University Press, Cambridge, 1958.
- [19] W.V.O. Quine, *Set Theory and Its Logic*, Harvard University Press, Cambridge, 1963.

- [20] B. Russell, ‘On some difficulties in the theory of transfinite numbers and order types’, *Proc. of London Math. Soc.* (1906); reprinted in *Essays in Analysis*, ed., D. Lackey, Braziller, New York, 1973.
- [21] B. Russell, and A. Whitehead, *Principia Mathematica*, Cambridge University Press, 1910–1913.
- [22] E. Specker, ‘The axiom of choice in Quine’s New foundations for Mathematical Logic’, *Proc. of the Nat. Acad. of Sciences* 39 (1953), 972–975.
- [23] E. Specker, ‘Typical ambiguity’, in *Logic, Methodology and Philosophy of Science*, eds., E. Nagel *et al.*, Stanford University Press, 1962. pp. 116–124.
- [24] H. Wang, ‘A formal system of logic’, *Journal of Symbolic Logic* 15 (1950), 25–32.
- [25] H. Wang, *From Mathematics to Philosophy*, Humanities Press, New York, 1974.

cocchiar@indiana.edu
Department of Philosophy,
Sycamore Hall 026,
Indiana University,
Bloomington IN 47405
U.S.A.