Typical Ambiguity: Trying to Have Your Cake and Eat it too.

Solomon Feferman

Would ye both eat your cake and have your cake?
John Heywood, *Proverbs*[1]

Abstract. Ambiguity is a property of syntactic expressions which is ubiquitous in all informal languages–natural, scientific and mathematical; the efficient use of language depends to an exceptional extent on this feature. Disambiguation is the process of separating out the possible meanings of ambiguous expressions. Ambiguity is typical if the process of disambiguation can be carried out in some systematic way. Russell made use of typical ambiguity in the theory of types in order to combine the assurance of its (apparent) consistency ("having the cake") with the freedom of the informal untyped theory of classes and relations ("eating it too"). The paper begins with a brief tour of Russell's uses of typical ambiguity, including his treatment of the statement $Cls \in Cls$. This is generalized to a treatment in simple type theory of statements of the form $A \in B$ where A and B are class expressions for which A is prima facie of the same or higher type than B. In order to treat mathematically more interesting statements of self membership we then formulate a version of typical ambiguity for such statements in an extension of Zermelo-Fraenkel set theory. Specific attention is given to how the"naive" theory of categories can thereby be accounted for.

1. Ambiguity, disambiguation and typical ambiguity

Ambiguity is a property of syntactic expressions which is ubiquitous in all informal (if not formal) languages–natural, scientific and mathematical; the efficient use of language depends to an exceptional extent on this feature. *Disambiguation* is the process of separating out the possible meanings of ambiguous expressions. Ambiguity is *typical*, or systematic, if the process of disambiguation can

[1]Ascribed to John Heywood (c. 1497 – c. 1580) in [3] at Heywood 13.

be carried out in some uniform way. In natural languages this is determined by specification of context. For example, the words *'I', 'you', 'here', 'now', 'this', 'my'*, and *'your'* used in daily language have meanings which vary with the context of utterance in a uniform way. Other words with more than one meaning such as 'bank', 'duck', and 'visiting', such as in the utterances 'He sat by the bank,' 'I saw her duck,' and 'Visiting relatives can be boring', can also be disambiguated from the context but not in any uniform way.[2]

Ambiguity in mathematical language involving such words as 'member,' 'union,' and 'complement,' is typical in that they are disambiguated in a uniform way relative to a context given by an underlying universe; similarly for the 'identity map.' In order to disambiguate arithmetic and algebraic symbols such as '0,' '1,' '+,' '×,' and '≤,' we need to be told just how these are to be interpreted in a given discussion.

I have not done a serious search of the use of the idea of typical or systematic ambiguity in mathematics and logic. The first employment of it that I've found (though without its being named in that way) was in Bertrand Russell's development of the theory of classes, relations and cardinal and ordinal numbers in his 1908 paper "Mathematical logic as based on the theory of types" [18]. How he used it there will be described in the next section. The idea is of course repeated and expanded in *Principia Mathematica*, beginning in Vol. I, from *20 on. But I did not find the words 'typical ambiguity' there either, except that in *65 there is talk of typically ambiguous symbols. Later uses by others are due, to begin with, to Quine ([16] and [17]) in connection with his system NF; the actual words 'typical ambiguity' are found in the second of these papers, pp. 132ff. The seminal work by Specker [19] aimed at proving the consistency of NF is entitled "Typical ambiguity"; that is tangentially related to what we are concerned with here in a way that will be briefly explained in the concluding section to this paper.

Russell used typical ambiguity in the theory of types to make sense of sentences like $Cls \in Cls$, which can be read as justifying talk in some sense of the class of all classes. In Section 3 this is generalized in a straightforward way to a treatment in simple type theory of statements of the form $A \in B$ where A and B are class expressions for which A is prima facie of the same or higher type than B. My main aim here is to extend that to mathematically interesting statements of the form $A \in B$ such as that the structure A we call the category of all categories is indeed a member of the class B of all categories. A few other test challenges of that sort are listed in Section 4. Then a version of typical ambiguity is formulated in an extension of ZF set theory in Section

[2]According to Godehard Link, "Shared opinion in the philosophy of language has it that in the case of indexicals like 'I','you', etc., it is not their meaning that varies with the context but rather their interpretation... This is considered to be different from lexical ambiguity ('bank'), syntactic ambiguity ('visiting relatives') or a combination of both ('I saw her duck')," [e-mail communication of 12 August 2003] I have received a similar comment from Thomas Wasow.

5 and applied to the test cases in Section 6. The final Section 7 concludes with a discussion of the problem of justifying problematic membership statements when read literally.

2. Russell's uses of typical ambiguity

The theory of types described in [18] is the ramified theory, but what is essential for his use of typical ambiguity there is best explained in terms of the simple theory of types (STT) with the Axiom of Infinity. The types are thus indexed by the natural numbers $0, 1, 2, \ldots$ with the objects of type 0 interpreted as the individuals (infinite in number) and for each n, the objects of type $n + 1$ interpreted as all classes of objects of type n. We use x, y, z, ... as variables for any type $n \neq 0$ and a matching list of variables X, Y, Z, ... of type $n + 1$.[3] Only atomic formulas of the form $u \in W$ where u is of some type n and W of the next type $n + 1$ are considered to be meaningful.

The first place that the issue of ambiguity comes up in [18] is on p. 251 (or p. 174 in its reprinting in [20]. For any type $n \neq 0$, he defines Cls[4] to be the class of all classes of objects of type $n - 1$ and then writes:

> ...the proposition '$Cls \in Cls$'...requires that 'Cls' should have a different meaning in the two places where it occurs. The symbol 'Cls' can only be used where it is unnecessary to know the type; it has an ambiguity which adjusts itself to circumstances.

A little further down he defines the empty class Λ and the universal class V and says that like Cls, these symbols are ambiguous "and only acquire a definite meaning when the type concerned is otherwise indicated." In our setting, these are particular objects of type n. Moving on from there Russell defines the Boolean operations $x \cup y, x \cap y$, and $-x$; this last is of course the complement relative to V. Later in the article he defines $Cl(x)$ to be the class of all subclasses of x; it is thus an object of type $n + 1$. Extensionally, Cls is the same as $Cl(V)$.

The theory of types was created in order to save the logistic program from inconsistency, in particular to avoid Russell's paradox. The use of variables $x, y, z, \ldots$ of an indefinite type and the notions just explained illustrates the use, when necessary, of typical ambiguity to fall back on the underlying type

[3] Russell uses α, β, γ, ... for class variables and x, y, z, ... for variables for their elements. Our choice of X, Y, Z, ... for class variables of type $n + 1$ has been made so that one can more easily compare typical ambiguity in type theory with its use in set theory that we take up in sec 5 below.

[4] In his [18], Russell wrote '*cls*' where I write '*Cls*'; the latter is used in *Principia Mathematica.*

structure for the security it provides ("having your cake") while regaining as much of the freedom of the informal theory of classes as possible ("eating it too").

Expanding the above form of STT to a theory of relations (as in Russell's theory) one can define the relation $x \sim y$ to hold when the classes x and y are in one-to-one correspondence. Then Russell defines $Nc(x)$ to be the class of all y such that $x \sim y$, i.e., $Nc(x)$ is the equivalence class of x under $\sim$; it is of type one higher than that of x. The class of cardinal numbers NC is then taken to be the class of all $Nc(x)$ for $x \in Cls$; it is of type $n+2$ for x of type n. The members 0, 1, 2, ... of NC are next introduced, respectively as $Nc(\Lambda)$, $Nc(\{0\}), Nc(\{0,1\})$, and so on. Of these definitions, Russell writes:

> It has to be observed...that 0 and 1 and all the other cardinals...are ambiguous symbols, like Cls, and have as many meanings as there are types. To begin with 0: the meaning of 0 depends upon that of Λ, and the meaning of Λ is different according to the type of which it is the null class. Thus there are as many 0's as there are types; and the same applies to all the other cardinals.

Russell's partial way out of this embarrassing situation is to note that if classes x and y are of different types, for example x of a type n and y of type $n+1$, then we can speak of x and y having the same cardinal number or of one having a larger cardinal number than the other by comparing y with the class of singletons $\{u\}$ for $u \in x$. But this still does not get out of the fact that one has a multiplicity of representatives of the cardinals and in particular of the natural numbers. To an extent, the use of typical ambiguity is a way of saving face in this respect.[5]

The general problem that concerns us here is how to interpret expressions of the form $A \in B$ which are of indefinite type, but where the prima facie type of A is greater than or equal to that of B. Russell himself signaled this issue when suggesting how to deal with the pseudo statement $Cls \in Cls$; one simply interprets the second occurrence of 'Cls' as being the class of all classes of type $n+1$, when the first occurrence is interpreted as the class of all classes of type n. A simpler statement which is meaningless on the strict account of type theory but which trades on the ambiguity of the symbols involved is

[5] Link has suggested that something like the treatment of the meaning of indexicals in natural language (as quoted in ftn.3 above) could be applied to the notion of number in the theory of types: one could say that the meaning of the number terms is always the same as defined via an expression with bound class variables but that only the "typical" context (namely a given type level) determines which classes are being bound. While I am not advancing any theory of meaning here, in a sense this is what is done in Sec. 3, where the meaning is determined by a formula φ. But the additional problem comes when meeting such examples as (1),(2) in section 2, so something more has to be done to take specify the interpretations of the terms in formulas of the form $A \in B$, if both A and B contain the same terms, e.g. $\{1,2\} \in 2$. This is what the disambiguation conventions are supposed to take care of.

$V \in V$, again legitimized by shifting the interpretation of the second 'V' one type higher.

There are other natural examples of this problem that Russell could have considered but did not. For example, the following statements seem reasonable (on the Axiom of Infinity):

(1) $$\text{Inf} \in \text{Inf},\ \text{Fin} \in \text{Inf},\ \text{Inf} \notin \text{Fin},\ \text{Fin} \notin \text{Fin},$$

where Fin, Inf are defined respectively to be the class of all finite classes and that of all infinite classes. So also are the statements:

(2) $$\{1\} \in 1,\ \{2\} \in 1,\ \{1\} \notin 2,\ \{1, 2\} \in 2.$$

3. Disambiguation of membership statements in type theory

Ambiguous expressions like Cls, Λ, V, 0, 1, ... can be assigned many types in STT; they are examples of expressions which are *stratified* in Quine's sense, i.e., result from erasing type indices from all the variables of an expression of STT (keeping variables of distinct types disjoint from each other). Given an expression S of STT, let $e(S)$ be the result of erasing all the type indices from variables of S. For A a stratified expression let type(A) be the least n for which there is an S of type n with $A = e(S)$; this is what we call the *prima facie type of* A. We shall confuse a stratified expression with the lowest expression of STT from which it results by erasing type distinctions. The disambiguation of expressions of the form $A \in B$ where the type of A is greater than or equal to that of B is first done here for the special case that they are of the same type and B is an expression of the form $\{x|\varphi(x)\}$, where φ is a stratified formula. Let φ^+ be stratified by replacing each variable in φ by the corresponding variable of next higher type, and then let B^+ be $\{X|\varphi^+(X)\}$. Then the *disambiguation convention* is simply:

(Dis 1) $$A \in B \text{ means } A \in B^+ \text{ when type}(A) = \text{type}(B).$$

Thus for B of the form $\{x|\varphi(x)\}$, $A \in B$ is equivalent to $\varphi^+(A)$. For example, the statements (1) above may be inferred as an application of (Dis 1). Similarly, if A is of prima facie type one higher than $B = \{x|\varphi(x)\}$, we agree to the following disambiguation convention:

(Dis 2) $$A \in B \text{ means } A \in B^{++} \text{ when type}(A) = \text{type}(B) + 1.$$

Then, for example, the statements (2) above may be inferred as an application of (Dis 2).

The ambiguity in the kinds of formulas $A \in B$ considered here is typical because disambiguation is systematic using the principles (Dis 1) and (Dis 2) depending on the type of A relative to that of B, and similarly when A is of still higher type. It is also typical because it doesn't really depend on the choice of type assignments on the basis of which the type of A is measured in comparison with that of B, as long as all types are shifted by the same amount. This is because of the straightforward:

Theorem 1. *If a sentence θ is provable in STT then so also is θ^+.*

It follows that if $A \in B$ is provable in STT where A, B are given by closed terms, then so also is $A^+ \in B^+$. More importantly, if $A \in B$ is provable then every property of elements of B also holds of A. This is formulated as:

Theorem 2. *(Transfer rule). For closed terms A, B, C with $B = \{x|\varphi(x)\}$ and $C = \{x|\psi(x)\}$ and type(A) = type(B), if $A \in B$ and $B \subseteq C$ are provable in STT then so also is $A \in C$.*

Proof. This is because $A \in B$ means $\varphi^+(A)$. Since $\forall x[\varphi(x) \to \psi(x)]$ is a theorem, so also is $\forall X[\varphi^+(X) \to \psi^+(X)]$, so $\psi^+(A)$ holds, i.e., $A \in C$ holds by the disambiguation convention. □

4. Some mathematical challenges.

All the preceding is quite obvious; moreover the applications are of limited mathematical interest, because STT doesn't lend itself to the flexible expression of mathematical properties in practice. Here are some statements of the form $A \in B$ with the prima facie type of A greater than or equal to that of B that are intuitively true in a naive theory of structures but cannot be verified directly in current systems of type theory or set theory. The aim is to make sense of them by some form of typical ambiguity.

Note that in the following we use (. , .) for the pairing operation, which is iterated to form n-tuples. We assume understood the mathematical notions involved (cf. [13] for the examples 4.2-4.4).

4.1 Let P be the class all partially ordered structures and let S be the substructure relation. Then $(P, S) \in P$.

4.2 Let *Set* be the category of all sets, *AbGrp* the category of all Abelian groups, *Top* the category of all topological spaces, etc., and let *CAT* be the class of all categories. Then, as should be, $Set \in CAT, AbGrp \in CAT, Top \in CAT$, etc.

4.3 Also $Cat \in CAT$ where $Cat = (CAT, FUNCT, \circ)$ is the *category of all categories*, whose objects are all categories and whose morphisms are the functors between categories, and $\circ$ is the partial operation of composition of functors.

4.4 If $A \in CAT$ and $B \in CAT$ then $B^A \in CAT$ where

$$B^A = (FUNCT(A, B), NAT(A, B), \circ),$$

whose objects are the class $FUNCT(A, B)$ of all functors from A to B and whose morphisms are the natural transformations between functors and where $\circ$ is the operation of composition of such transformations.

4.5 Let BA be the class of all Boolean algebras. Then $(\wp(V), \cup, \cap, -, \oslash, V) \in BA$ where V is the universal class and $\wp(V)$ is the class of all subclasses of V.[6]

5. Typical ambiguity in a system of set theory with universes

Let L be the language of ZF set theory, using variables x, y, z, ... for sets. Adjoin to L constants U_n for $n = 0, 1, 2, \ldots$ for an increasing sequence of *reflective universes;* the resulting language is denoted L$(U_{<\omega})$, where we use $U_{<\omega}$ to indicate the sequence of U_n for $n < \omega$. Each U_n is supposed to be a set which is reflective in the sense that–speaking model-theoretically–it forms an elementary substructure of $(V, \in)$ when the membership relation is restricted to U_n. As usual, to express this, we use the operation φ^a of forming the relativization of all quantifiers in the formula φ of L to a; we also write Rel(φ, a) for the resulting formula. The universes are also supposed to be *supertransitive*, i.e. transitive and closed under the power set operation $\wp$; for the latter it is sufficient to assume that universes are closed under subsets of members. The system of ZF, resp. ZFC, with universes satisfying these properties is denoted ZF$/U_{<\omega}$, resp. ZF$/U_{<\omega}$. More officially, the axioms are as follows.

Axioms of ZF$/U_{<\omega}$

I. All the axioms of ZF in L, and for each $n = 0, 1, 2, \ldots$:

II. $U_n \in U_{n+1}$

III. Trans(U_n).

IV. $\forall x \forall y [y \in U_n \wedge x \subseteq y \rightarrow x \in U_n]$.

[6]This example was suggested to me by Godehard Link.

V. For each L formula $\varphi(x_1, \ldots, x_k)$,
$\forall x_1 \ldots \forall x_k \{x_1, \ldots, x_k \in U_n \rightarrow [Rel(\varphi, U_n)(x_1, \ldots, x_k) \leftrightarrow \varphi(x_1, \ldots, x_k)]\}$.

The Axioms of ZF/$U_{<\omega}$ are obtained by adding the Axiom of Choice, AC.

Theorem 3. *ZF/$U_{<\omega}$ is a conservative extension of ZF.*

Proof. This is by a straightforward extension of the result of Montague and Vaught [14] which in turn modifies Levy's well known reflection principle argument; see also ([5]: sec. 2). We give here a brief sketch of the ideas involved. Note that for conservation we need only show how, given any finite set of axioms of ZF/$U_{<\omega}$ involving only symbols U_i for $i = 0, \ldots, n$ for some n, to define sets in ZF satisfying the given axioms for these universes. Working informally in ZF, let V_α be the αth set in the cumulative hierarchy, i.e. for each α, V_α is the union of $\wp(V_\beta)$ for all $\beta < \alpha$. For any set x, the rank of x, $\rho(x)$, is the unique α such that $x \in V_{\alpha+1} - V_\alpha$. One associates with each existential formula $(\exists x)\Psi(x, y_1, \ldots, y_m)$ of L an m-ary function F_Ψ whose value $F_\Psi(y_1, \ldots, y_m)$ for each $y_1, \ldots, y_m$ is the set of all x of least rank such that $\Psi(x, y_1, \ldots, y_m)$ holds. Thus

$$(\exists x)\Psi(x, y_1, \ldots, y_m) \rightarrow (\exists x \in F_\Psi(y_1, \ldots, y_m))\Psi(x, y_1, \ldots, y_m).$$

The F_Ψ act like Skolem functions, but without needing the Axiom of Choice to pick a specific value of witness for the existential quantifier. By the *rank-hull* of a set b we mean the least V_α such that $b \subseteq V_\alpha$. For any finite set Q of formulas $(\exists x)\Psi(x, y_1, \ldots, y_m)$, by the *Skolem-rank-hull* of a set b relative to Q, in symbols $H_Q(b)$, we mean the least V_α such that $b \subseteq V_\alpha$ and such that for each $(\exists x)\Psi(x, y_1, \ldots, y_m)$ in Q and each $y_1, \ldots, y_m \in V_\alpha, F_\Psi(y_1, \ldots, y_m) \subseteq V_\alpha$. $H_Q(b)$ is obtained as the union of bj for $j < \omega$, where $b_0 = b$ and each b_{j+1} is the rank hull of $[b_j$ union all $F_\Psi(y_1, \ldots, y_m)$ for all formulas $(\exists x)\Psi(x, y_1, \ldots, y_m)$ in Q and all $y_1, \ldots, y_m \in b_j]$. For simplicity, assume the formulas of L are generated from atomic formulas $x \in y$ and $x = y$ by $\neg, \wedge$ and $\exists$. Now, given the finite set of axioms of ZF/$U_{<\omega}$ to be modeled in ZF, let S be the closure under subformulas of all the L formulas in the given set and let Q consist of all existentially quantified formulas $(\exists x)\Psi$ in S. Then it is proved by formula induction that for each set b and each φ in S,

$$\forall x_1 \ldots \forall x_k \{x_1 \ldots x_k \in H_Q(b) \rightarrow [Rel(\varphi, H_Q(b))(x_1 \ldots x_k) \leftrightarrow \varphi(x_1 \ldots x_k)]$$

Since also each $H_Q(b)$ is supertransitive by construction, we can define U_i for $i = 0, \ldots, n$ by taking $U_0 = H_Q(0)$ and for each $i < n, U_{i+1} = H_Q(U_i \cup \{U_i\})$. □

NB. If θ is any sentence of L then adding θ as an axiom maintains conservativity by this theorem.

Corollary. *ZFC/$U_{<\omega}$ is a conservative extension of ZFC.*

Note that each reflective universe U_n satisfies all the axioms of ZF by taking closed φ in Axiom V to be any one of these axioms. It follows by supertransitivity that each universe contains the empty set 0 and the set ω of finite ordinals, is closed under unordered pair, union and power set; moreover, these are absolute, i.e. have the same values in the universe as in the universe of all sets. Moreover, each universe U_n is closed under the Separation Axiom and Replacement Axiom schemes. The latter can be thought of as follows: if f is any function in the set-theoretic sense of the word that is *defined in* L (with respect to any given parameters in U_n) and if $a \in U_n$ and $f : a \to U_n$, then $\{f(x) : x \in a\} \in U_n$. If one wishes to drop here the assumption that f is defined, allowing it to be *any function*, then $\rho(U_n)$ would have to be a strongly inaccessible cardinal, and assumption of that would no longer hold conservatively over ZF; instead we would have to strengthen ZF by the assumption of the existence of infinitely many strongly inaccessible cardinals. This might be needed for some applications (see the next section), but is not assumed here.

We now consider to what extent application of typical ambiguity in the system ZF/$U_{<\omega}$ (or the same augmented by AC as dictated by specific needs) can be used to meet the mathematical challenges of the preceding sec.4.[7] Looked at informally, what we have to deal with to begin with in the examples 4.1-4.3 are membership relations of the form $A \in B$, where B is a class and A is a (possibly) many-sorted relational structure each of whose domains and relations are classes. In all the cases considered, B is given as $\{x|\varphi(x)\}$ for some L formula φ; the classes which are the constituents of A are also defined in L. We shall relativize the concepts involved in A and B to universes. By a *typical reflective universe* U, we mean *any* U_n; then by the *next universe* U^+ we mean U_{n+1}. The first step in interpreting the problematic membership statements relative to any such universe U is to *identify classes with the corresponding subsets of* U. The second step is to *identify sets with the members of* U. Note that since the classes making up the structure A are subsets of U, we have $A \in U^+$. Now, to make sense of $A \in B$, the third step is to *re-identify* B *with the class of all sets in* U^+ *that satisfy the definition of* B; call this B^+. More precisely, if B is given formally as $\{x|\varphi(x)\}$, its first identification is with $\{x \in U|\varphi^U(x)\}$ which is the same as $\{x \in U|\varphi(x)\}$; then, the re-interpretation B^+ of B is simply

[7]Something like this was suggested as follows by Kurt Gödel in a letter to Paul Bernays dated 1 January 1963: "I find it interesting that you speak... of the 'newer abstract disciplines of mathematics' as something lying outside of set theory. I conjecture that you are thereby alluding to the concept of category and to the self-applicability of categories. But it seems to me that all of this is contained within a set theory with a finitely iterated notion of class, where reflexivity results automatically through a 'typical ambiguity' of statements." (The source for the German original of this letter is in the Bernays archives at the ETH in Zürich; the full letter is to be found, along with much else between Gödel and Bernays, in the collection of correspondence in [9].)

defined to be $\{x \in U^+ | \varphi(x)\}$. This leads to the disambiguation convention:

(Dis 3) $A \in B$ means $A \in B^+$.

To take the simplest example, 4.1, P is defined to be the class of structures (x, y) where $y \subseteq x^2$ is a partial ordering of x, and S is defined to be the substructure relation $(x, y) \subseteq (z, w)$ on P which holds just in case $x \subseteq z$ and $y = w \cap x^2$. With the variables x, y, z, w taken to range over the sets in any typical reflective universe U, we have that P and S are subsets of U. According to the disambiguation principle (Dis 3), $(P, S) \in P$ means that $(P, S) \in P^+$, where P^+ consists of all those sets (X, Y) with X, Y in U^+ such that $Y \subseteq X^2$ and Y is a partial ordering of X. It is readily verified that in this case, we do indeed have $(P, S) \in P^+$, so it makes sense by the disambiguation principle to assert that $(P, S) \in P$. In words: the class of all partially ordered structures (in a typical universe) together with the substructure relation (in that universe) forms a partially ordered structure (in the next universe).

There is more to be said about the above disambiguation principle, again illustrated by reference to 4.1. By the reflection axiom V of ZF/$U_{<\omega}$, P and P^+ share exactly the same properties expressed in the language L of set theory, and *each* relativized to any typical reflective universe can be considered as a *surrogate* of the class of all sets (x, y) which form partial ordering structures. Moreover, consider any property $\Psi(x, y)$ formulated in L which is proved to be true of all such (x, y). Let C be defined as the class of all (x, y) such that $\Psi(x, y)$ holds. This definition is taken to be ambiguous, i.e. it can be thought of as pertaining to the class of all sets, or relativized to any universe. Our assumption is that $P \subseteq C$ has been proved in ZF, i.e. when we relativize to the class of all sets. It follows that $P^+ \subseteq C^+$ when P is interpreted as the class of all partially ordered structures in any typical reflective universe U. From the disambiguation principle, it follows that $(P, S) \in C$. In other words, (P, S) *shares all the properties that are verified of all partially ordered structures*; this is an example of the *transfer rule* in this framework. More generally, we have:

Theorem 4. *(Transfer rule). If $A \in B$ and $B \subseteq C$ are provable in ZF where B, C are class abstracts then $A \in C$ is provable there too.*

Proof. For B given formally as $\{x|\varphi(x)\}$ and C as $\{x|\Psi(x)\}, B \subseteq C$ is read as $\forall x[\varphi(x) \rightarrow \Psi(x)]$, and what it means for $B \subseteq C$ to be provable in ZF is that $\forall x[\varphi(x) \rightarrow \Psi(x)]$ is provable in ZF. Then by the reflection axiom V, given any universe $U, (\forall x \in U^+)[Rel(\varphi, U^+)(x) \rightarrow Rel(\Psi, U^+)(x)]$, i.e., $B^+ \subseteq C^+$. From provability of $A \in B$, which means that $A \in B^+$ by (Dis 3), it follows that $A \in C^+$, i.e. $A \in C$, again by the disambiguation convention. □

To look at things more generally, consider a type structure over sets, where sets are considered to be at type level 0, classes of sets at type level 1, classes of classes of sets of type level 2, and so on. We regard the type level of a structure A to be the same as the maximum type level of the domains, relations, operations and individuals that make up that structure. Thus in the examples 4.1-4.3 of problematic membership statements of the form $A \in B$ the type level of both A and B is 1. But in the example 4.5, the structure $A = (\wp(V), \cup, \cap, -, \emptyset, V)$ is of type level 2 since the class V of all sets is of type level 1 and $\wp(V)$, the class of all subclasses of V, is of type level 2. On the other hand the type level of B–in this case the class BA of all sets which are Boolean algebras–is still 1. But unlike the case of type theory, where to disambiguate $A \in B$ we had to go to $A \in B^{++}$ when the type level of A is one higher than that of B, we can still disambiguate by (Dis 3). For, when we interpret the sets as ranging over any typical universe U, A is interpreted as being the structure $(\wp(U), \cup, \cap, -, \emptyset, U)$ and since U and $\wp(U)$ belong to U^+, the structure A in this case also belongs to U^+, and in fact is a member of the subset BA^+ of U^+ consisting of all structures in U^+ which satisfy the conditions to be a Boolean algebra. Thus it makes sense by (Dis 3) to say that $A \in B$ holds in this case. The advantage over ordinary type theory given to us by the set-theoretical framework is that the type level of U^+ over U measured in terms of the cumulative hierarchy is infinite. A precise explanation of how (Dis 3) may be applied more generally would require the introduction of an extension L^+ of L by variables of classes, which are assumed (unlike Gödel-Bernays or Morse-Kelley theories) to satisfy all the axioms of ZF. Then for any application, the ZF objects are interpreted as ranging over a universe U and the classes are interpreted as ranging over the next universe U^+. For the purposes here, it is sufficient to see how (Dis 3) works in a few specific cases, like that of 4.5 just discussed, or (4.4) to be treated in the next section.

6. Category theory in ZF/U$_{<\omega}$

The reader is assumed to be familiar with the basic notions of category theory. A standard reference is [13]. We can take categories to be structures of the form $A = (O, M, C)$ where O is the class of objects of A, M its class of morphisms and C the composition of morphisms in $A; C$ is a partial operation on M^2 to M considered as a three-placed relation on M. Actually, we can identify O with the class of identity morphisms, suitably defined, and the domain and codomain of a morphism with its left and right identities, so it is sufficient to consider structures of the form (M, C). Alternatively, and more intuitively, we can take categories to be structures of the form $A = (O, M, C, D_0, D_1, I)$ where C is as before, the D_0 and D_i are maps from M to O giving the domain and codomain respectively, and I is a map from O to M. As usual we write $f : x \to y$ or $x \to^f y$ in A when $f \in M, D_0(f) = x$ and $D_1(f) = y$; id_x

for $I(x)$; and $fg = h$ when $(f, g, h) \in C$. When it is necessary to compare objects from one category with another we subscript the terms of A by 'A' as $A = (O_A, M_A, C_A, D_{0A}, D_{1A}, I_A)$. A is a *large category* if O_A and M_A are both classes. A is said to be *locally small* if for any objects x, y of A the class $H_A(x, y) = \{f | f \in M_A \wedge f : x \to y\}$ is a set; it is *small* if A itself is a set.

The following are standard examples of locally small categories.

(The category of all sets) *Set* is the category whose objects are just the sets and whose morphisms are all triples $f = (u, x, y)$ where u is a function (in the usual set-theoretical sense) whose domain is x and range is contained in y. Then $D_0(u, x, y) = x$ and $D_1(u, x, y) = y$. The composition fg is defined for $f = (u, x, y)$ and $g = (v, z, w)$ just in case $y = z$, in which case $fg = (u; v, x, w)$ where $u; v$ is the relational composition of u and v. Finally $I(x)$ is taken to be (u, x, x) for each x, where u is the identity function from x to x.

(The category of Abelian groups) *AbGrp* is the category whose objects are the sets which are Abelian groups and whose morphisms are group homomorphisms with specified domain and codomain. This is then spelled out as in 6.1.

(The category of all topological spaces) *Top* is the category whose objects are the members of U that are topological spaces, treated similarly.

(The category of all categories) *Cat* is the category whose objects are the sets (o, m, c, d_0, d_1, i) that satisfy the conditions to be a category, and whose morphisms are the functors between any two such categories. *Cat* itself has the form $(CAT, FUNCT, \ldots)$ where $CAT = O_{Cat}$ and $FUNCT = M_{Cat}$.

Now by (Dis 3) we can make sense of the statements in 4.2 and 4.3 that $Set \in CAT, AbGrp \in CAT, Top \in Cat$ and even $Cat \in CAT$, i.e., $(CAT, FUNCT, \ldots) \in CAT$. Disambiguated, these statements mean that relative to any typical reflective universe U, each of $Set, AbGrp, Top$, and Cat is an element of CAT^+, the class of all sets $A = (O, M, \ldots)$ in the next universe U^+ which satisfy the conditions to be a category. Furthermore, the transfer theorem assures us that *any properties that apply to all members of CAT, i.e. the class of all <u>small</u> categories, also hold for the <u>large</u> categories Set, AbGrp, Top, Cat, etc.*

(The category of all functors between two given categories) When we turn to making sense of the statement 4.4 that $B^A \in CAT$ where A and B are large categories we meet the problem that the prima-facie type level of B^A is higher than that of both A and B. In MacLane's terminology, B^A is beyond large, sometimes called *superlarge*. But from the point of view of the method of disambiguation proposed in the preceding section, the issue here is no different from that met with the example 4.5 concerning the Boolean algebra on $\wp(V)$. But there are mathematical aspects of example 4.4 that are worth a closer look. Consider the usual set-theoretical definition of B^A: its objects are $FUNCT(A, B)$, i.e., the class of all functors $F : A \to B$, and its morphisms are natural transformations $\eta : F \to G$ between such functors. We can take functors to be classes which are functions in the usual set-theoretical sense, so for each $x \in O_A, F(x) \in O_B$; as usual, to be a functor, F is required to preserve composition and identity morphisms, i.e., for $f, g \in M_A$ and $x \in O_A, F(fg) = F(f)F(g)$ in B and $F(i_x) = i_{F(x)}$. By a natural transformation $\eta : F \to G$ between two such functors $F : A \to B$ and $G : A \to B$ is meant a map from O_A to M_B such that the following diagram is commutative for any $x, y \in O_A$, whenever $x \to^u y$ in A:

$$\begin{array}{ccc} F(x) & \xrightarrow{F(u)} & F(y) \\ {\scriptstyle \eta(x)}\downarrow & & \downarrow{\scriptstyle \eta(y)} \\ G(x) & \xrightarrow{G(u)} & G(y) \end{array}$$

Composition of natural transformations is defined in an obvious way. Though, as noted above, B^A is no longer a large category when A and B are large, we can still make use of the disambiguation convention (Dis 3) as at the end of the preceding section to make sense of the statement that it is a member of CAT, since, according to it, this simply means that relative to any typical universe $U, B^A \in CAT^+$. [8] There is no problem with this, since now with the O_A, M_A, O_B, and M_B regarded as subsets of a given universe U, the sets O_C and M_C belong by the above definition to U^+ for $C = B^A$. Again, the transfer theorem assures us *that any property of all small categories also applies to the superlarge category* B^A.

7. Discussion: Axiomatic foundations of category theory

One of the usual foundations of category theory, due to MacLane [11], takes as its setting the language of the Bernays-Gödel theory of sets and classes; this allows us to talk about two kinds of categories, those that are *small*, i.e. are sets, and those that are *large*, i.e. are classes which are not sets. In such a foundation, when A and B are both large categories, such as $A = AbGrp$ and $B = Set$, there is no place to locate B^A, since it is now of a type higher than the classes in the sense of the language of BG. Thus one often is forced to restrict this construction to the case that A is small, so that the functors from A to B are all sets, and B^A is at most a large category in the BG sense. The preceding shows that that problem is not met in the ZF/$U_{<\omega}$ setting presented here. Thus, for example, we can state the Yoneda Lemma as a natural equivalence between any locally bounded (i.e., locally small) category A and the category $Set^{A'}$, where A' is the opposite category to A.

[8] This is in accord with Gödel's idea quoted in footnote 8.

Now it is to be noticed that we never needed more than two universes U and U^+ to take care of disambiguation in the various examples 4.1-4.5. Thus, we could just as well have restricted the theory to the part of ZF/$U_{<\omega}$ that concerns only U_0 and U_1. One could go even farther, by dealing with *only one reflective universe* U, of which (axiomatically) the full universe V is taken to be an elementary extension, thus with U being treated like U_0 and V like U_1; call this form of the theory ZF/U_0.[9] Large categories have as surrogates subcategories of U_0, and functor categories for them simply sit as sets in V. On the face of it, this would seem to make it similar to the suggestion of MacLane [12] according to which one universe suffices. But there is an essential difference here in the further use of the reflection Axiom V to insure that there is nothing special about the choice of the universe U_0; with respect to properties formulated in set-theoretical terms, it is indistinguishable from the full universe. Thus any property established for subcategories of U_0 holds of all categories, in particular the category Cat and the functor categories of 4.4.

Actually, the idea of using a theory like ZF/U_0 or ZFC/$U_{<\omega}$ as a foundational framework for category theory is an old one that I first elaborated in the paper "Set-theoretical foundations of category theory" (Feferman 1969).[10] What ZF/$U_{<\omega}$, resp. ZFC/$U_{<\omega}$, provides as an advantage is the ability to explain more directly the disambiguation relative to any typical reflective universe in terms of the next universe. The adequacy of ZFC/U_0 as a framework for working category theory was considered with respect to some prominent test cases, including: (i) Yoneda Lemma, (ii) Freyd Adjoint Functor Theorem, and (iii) the functors Ext_n in homological algebra. These, and others, should be re-considered as test cases for a more flexible development in the system ZFC/$U_{<\omega}$.

In some cases we may need somewhat stronger hypotheses; for example (as explained in [5], it appears that the Kan Extension theorem requires the rank of the typical universe U being considered to be a strongly inaccessible cardinal. To deal with such cases we shall need to add as an assumption to our base system of set theory that beyond any ordinal there is a strongly inaccessible cardinal. (Such an assumption is considered to be innocuous by working set theorists.) A proposed foundation for category theory ascribed to

[9]That was the form in which typical ambiguity in set theory was presented in a draft of this paper. The reasons for the shift to a theory with many reflective universes are given below.

[10]A similar proposal has more recently been made by F.A. Muller [15], taking an extension of the theory of sets and classes due to Ackermann, as the basic framework; this system is known to be interpretable in ZF/U_0 by interpreting the sets to be the members of U_0 and the classes to be the sets in V. Muller incorrectly asserts that my paper, Feferman (1969) required the assumption of inaccessible cardinals. For a critical review by Andreas Blass of Müller (2001) and elaboration of its relation to my earlier work, see http://www.ams.org/mathscinet/ review 2002k:03008.

Grothendieck makes use of this assumption; by a universe on that approach is meant any V_α for α a strongly inaccessible cardinal. What that foundation does not explain is why such universes may be considered typical. That is what the reflection Axiom V adds to the Grothendieck approach. But it is useful to see when the much weaker system ZFC/$U_{<\omega}$, without the assumption of any inaccessibles, suffices for various parts of the development of category theory; in fact, that assumption seems to be rarely needed.

Another, general, reason for dealing here with the system with a sequence of reflective universes rather than a single one has been to show how the handling of typical ambiguity in set theory parallels, in part, that in the theory of types as dealt with above in Sec. 3.

There are other candidates for foundational frameworks for category theory employing typical ambiguity which would seem to have some advantages over ZF/$U_{<\omega}$ with or without AC, especially with respect to the treatment of functions. One is a form of operational set theory, in which the system is extended further by variables for a partial combinatory algebra over the universe.[11] Another is to use one of the systems of Explicit Mathematics introduced in [7] and studied in a number of publications since then. There the operational structure in the form of a partial combinatory algebra on the universe of all individuals is taken as basic. For a system of that kind to work we would have to extend the formalism by symbols for typical reflective universes with the appropriate axioms. Both treatments allow B^A to be interpreted directly as a class of type level 1 rather than type level 2. I have done some experimentation with both of these approaches, but have not yet brought the work to a definitive form.

8. Trying to have your cake and eat it too: naive category theory.

Minimally, what one is after is to have a demonstrably consistent foundational framework T for mathematically interesting cases of self-membership such as provided by category theory. But, more than consistency, one would want (as with ZF/$U_{<\omega}$) conservativity over an accepted framework (in that case over ZF); all that is done here, though not without some clumsiness in the applications. Ideally, what one is after–and that is *not* done here–is to provide a framework meeting these criteria in which such objects as CAT exist

[11] See http://math.stanford.edu/ Feferman/papers/OperationalST-I.pdf for a draft article featuring such a system. (NB. Theorem 4(i) p.5 needs correction.) Incidentally, Vidhyanath Rao has informed me that there are some theorems which require global choice for their proof. The system of operational set theory referred to here incorporates a global choice operator.

in the universe of discourse and which are such that $(CAT, \ldots) \in CAT$ is *literally true*, and *not reinterpreted* according to context. For a discussion of the desiderata for that kind of foundation, see my paper "Categorical foundations and foundations of category theory" [8]. We have yet to obtain a satisfactory solution of the problem posed by naive category theory in that form.

At first sight, one direction in which such a solution might be sought is to make use of theories like that of Aczel [1] as exposited in [4], in which the Foundation Axiom is replaced by the Anti-Foundation Axiom AFA. That certainly allows many examples of self membership such as $a \in a$ and $(a, b) \in a$, and so on, with interesting applications when suitably elaborated. In Ch. 20 of that book there is a proposed extension SEC_0 of the theory of sets in which class variables are adjoined, that also allows some cases of self-membership between classes such as $A \in A$ and $(A, B) \in A$; *grosso modo*, such statements are needed to deal with the kinds of mathematical applications from category theory considered here. Some of the above desiderata are met by that system, namely that of consistency relative to ZFC, of which it is an extension. But, as the authors themselves point out, the mathematical usefulness of SEC_0 in general remains to be established.[12]

Another direction in which such a solution might be sought is via some form of stratified theories like Quine's NF. Though that is still not known to be consistent, the system NFU with urelements allowed was shown to be consistent by Jensen [10], and he also showed how it can be beefed up to include, conservatively, ZFC. However, the formalism of NFU is not as it stands suitable to define $\{x|x$ is a relational structure of signature $\sigma\}$ where σ is any specified signature. For example, if x is to be of the form (y, z) with $z \subseteq y^2$ we need to assign to the elements of z, which are ordered pairs, the same type level as the elements of y, and thus ordered pairs of elements of a set need to be assigned the same type level as the elements of that set. None of the usual definitions of ordered pair in NFU works to do that. However, one can formulate a simple extension NFUP of NFU in which pairing is taken as a basic operation, and stratification is modified so as to allow us to assign the same type to an ordered pair as to its terms. By an adaptation of Jensen's proof, one can establish the consistency of NFUP and again one can beef it up to obtain a version of it conservative over ZFC. I carried this out in an unpublished MS, "Some formal systems for the unlimited theory of structures and categories" (abstract in [6]).[13] This system indeed serves to literally verify examples like 4.1-4.4. But it has other defects as a proposed foundation of naive category theory. One is that there is no way that one can establish existence

[12] According to one of the referees, the work of Barwise and Moss on SEC_0 has been continued by Alexandru Baltag in his doctoral thesis, and one publication resulting from that is [2]. I do not know how relevant it may be to the questions dealt with here.

[13] The proof makes use of the existence of two inaccessible cardinals, an assumption which, as remarked above, is regarded as innocuous by working set-theorists.

of the cartesian product $\Pi x_i (i \in I)$ of a collection $\{x_i | i \in I\}$ of sets, since the collection must be given by a function g with $g(i) = x_i$ for each $i \in I$, and the elements of the cartesian product f must be of the form $(i, f(i))$ with $f(i) \in g(i)$. Thus there is no stratified type assignment for pairing which allows one to deal with both f and g simultaneously. On the other hand, there is no obvious way to obtain a consistent extension of NFU allowing stratification of pairs where the terms of a pair are of prima facie mixed type.

As the title of Specker [19] attests, there is of course a non-trivial connection of the consistency problem for stratified systems with typical ambiguity. Specker considered one form of typical ambiguity in STT to be given by the scheme $(\varphi \leftrightarrow \varphi^+)$ for all sentences φ of type theory; he showed that NF is consistent just in case STT is consistent when augmented by this scheme. One way to insure that would be to seek models of type theory in which there is a shifting endomorphism which takes each type level to its successor level, or a model of STT allowing negative types in which there is a type shifting automorphism. Jensen succeeded in using Specker's idea in his proof of consistency of NFU, by combining it with the Ehrenfeucht-Mostowski theorem on the existence of models with many automorphisms. But even if the consistency of NF were established via Specker's theorem or by some alternative approach, NF would not, as it stands, provide the interpretation of structural notions needed to insure literal self-membership in the sense of such examples as 4.1-4.4, as well as to satisfy the additional criteria of [8] for ordinary mathematical constructions such as that for product above.

Acknowledgement. Invited lecture for the conference, One Hundred Years of Russell's Paradox, University of Munich, June 2-5, 2001. I wish to thank Godehard Link for his exceptional work in organizing this conference and for his personal assistance in connection with it. I have received useful comments on this paper from Tim Fernando, Godehard Link, Karl-Georg Niebergall and two anonymous referees.

References

[1] Aczel, Peter: 1988. *Non-Well-Founded Sets*. Stanford: CSLI Publications no. 14.

[2] Baltag, Alexandru: 1998. STS: A structural theory of sets. *Advances in Modal Logic* 2:1–34.

[3] Bartlett, John 1980: *Familiar Quotations*, Fifteenth edition. Edited by E. M. Beck. Boston: Little, Brown and Co.

[4] Barwise, Jon and Larry Moss: 1996. *Vicious Circles*. Stanford: CSLI Publications no.60.

[5] Feferman, Solomon: 1969. Set-theoretical foundations of category theory (with an appendix by G. Kreisel). In: M. Barr *et al.* (eds.), *Reports of the Midwest Category Seminar III*, Lecture Notes in Mathematics 106, 201–247.

[6] Feferman, Solomon: 1974. Some formal systems for the unlimited theory of structures and categories (abstract). *J. Symbolic Logic* 39: 374–375.

[7] Feferman, Solomon: 1975. A language and axioms for explicit mathematics. In: J. Crossley (ed.), *Algebra and Logic*, Lecture Notes in Mathematics 450, 87–139.

[8] Feferman, Solomon: 1977. Categorical foundations and foundations of category theory. In: R. E. Butts and J. Hintikka (eds.), *Logic, Foundations of Mathematics and Computability Theory*, Vol.1, Dordrecht: Reidel, 149–169.

[9] Gödel, Kurt: 2003. *Collected Works, Volume IV. Correspondence A-G.* Edited by S. Feferman *et al.*, Oxford: Oxford Univ. Press.

[10] Jensen, Ronald: 1969. On the consistency of a slight (?) modification of Quine's *New Foundations*. In: D. Davidson and J. Hintikka (eds.), *Words and Objections: Essays on the Work of W.V.O. Quine*, Dordrecht: Reidel, 278–291.

[11] MacLane, Saunders: 1961. Locally small categories and the foundations of mathematics. In: *Infinitistic Methods*, Oxford: Pergamon Press, 25–43.

[12] MacLane, Saunders: 1969. One universe as a foundation for category theory. In. M. Barr *et al.* (eds.), *Reports of the Midwest Category Seminar III*, Lecture Notes in Mathematics, 106: 192–200.

[13] MacLane, Saunders: 1971. *Categories for the Working Mathematician.* Berlin: Springer.

[14] Montague, Richard and Robert L. Vaught: 1959. Natural models of set theories. *Fundamenta Mathematicae* 47: 219–242.

[15] Muller, Frederik A.: 2001. Sets, classes, and categories. *British J. Philosophy of Science* 52: 539–573.

[16] Quine, Willard V.O.: 1937. New foundations for mathematical logic. *American Mathematical Monthly* 44: 70–80.

[17] Quine, Willard V.O.: 1938. On the theory of types. *J. of Symbolic Logic* 3: 125–139.

[18] Russell, Bertrand: 1908. Mathematical logic as based on the theory of types. *American Mathematical Monthly* 30: 222–262. Reprinted in ([20]: 150–182).

[19] Specker, Ernst: 1962. Typical ambiguity. In: E. Nagel *et al.*(eds.), *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress*, Stanford: Stanford Univ. Press, 116–124.

[20] van Heijenoort, Jean (ed.): 1967. *From Frege to Gödel. A Source Book in Mathematical Logic*, 1879-1931. Cambridge, MA: Harvard Univ. Press.

[21] Whitehead, Alfred N. and Bertrand Russell: 1925. *Principia Mathematica*, Vol. I, second edition. Cambridge:Cambridge Univ. Press.

Department of Mathematics
Stanford University
Stanford CA 94305-2125 New York, NY 10003

Email: sf@csli.stanford.edu