Foundations of Mathematics in Polymorphic Type Theory

M. Randall Holmes

September 2, 2011

1 Introduction

This essay was inspired by conversations with mathematicians who maintain in all seriousness that there is something canonical about foundations of mathematics in *ZFC* (Zermelo-Fraenkel set theory with the axiom of choice). (These took place on the FOM (Foundations of Mathematics) mailing list [7].) One of the parties to this conversation ventured to *define mathematics* in terms of *ZFC*. The best way to disprove such claims is to present foundations for mathematics which do not depend on *ZFC*. This is what is done here.

The approach to foundations presented here is formalized in *NFU*, a theory proposed by Jensen in [17] as a "slight" modification of Quine's set theory "New Foundations" of [22] (hereinafter *NF*). Actually, by *NFU* we will always mean at least *NFU* + Infinity + Choice. We say that the approach is formalized in *NFU*; this does not mean that it is completely embodied in *NFU*. Just as the underlying intuitive picture of foundations which motivates *ZFC* suggests further extensions of *ZFC* (with inaccessible cardinals and so forth), so the underlying view of foundations which motivates *NFU* suggests extensions of the formal theory. Some of these extensions will be discussed here.

A common criticism of Quine's "New Foundations" is that it is not motivated by any underlying picture of the world of sets; that it is merely a modification of the theory of types via a "syntactical trick". This may be true of Quine's original development of *NF*, but we will present an intuitive picture of the world of sets as well as a formal system in this essay.

In [17], Jensen demonstrated the consistency of *NFU* + Infinity + Choice, and Jensen and later workers have described models of *NFU*. It is not hard to develop a clear mental picture of what a model of *NFU* is like. This does not fulfil our program of presenting *NFU* as an autonomous approach to the foundations of mathematics, because it can be objected (with considerable justice) that the whole warrant for these proceedings rests on a consistency proof carried out in *ZFC*, and that the models one can contemplate are structures in the world of *ZFC*.

To achieve the philosophical program of presenting *NFU* as an autonomous foundational approach, it is necessary to proceed more carefully. Our development will present foundations in *NFU* as a natural revision of foundations in the Theory of Types of Russell (as simplified by Ramsey, and also modified by a weakening of extensionality to allow urelements in each type). We will be interested at each step not only in mathematical soundness but in intuitive appeal of the constructions.

It is not the purpose of this paper to suggest that *ZFC* foundations are unsatisfactory or should be replaced by *NFU* foundations. It is not even a purpose of the paper to suggest that *NFU* foundations cannot be presented as a revision of *ZFC* foundations themselves (they can, and we indicate in the section titled "Mutual Reflections" how to do this, though we think that a development of *NFU* via type theory is more natural, especially since one needs to understand type theory anyway to see the point). No change in the practice of mathematics is being suggested here; our aim is to suggest that other approaches to the foundations of mathematics than the one now in fashion could have been taken, by exhibiting a different approach which is technically feasible and has a natural motivation of its own.

This paper is also influenced (though not visibly) by the fact that the author has implemented a theorem prover which uses a version of this approach to mathematics as its built-in logic. Experience with such a system provides evidence for the adequacy of the foundational view built into it (and, we must add, for problems with that foundational view). For information about the theorem prover, see [15]; for a technical treatment of its mathematical foundations, see [16].

2 The Theory of Types

We present a version of the Theory of Types of Russell, as simplified by Ramsey. We will refer to the theory we develop here as TTU (the U is for "urelements", as we weaken extensionality).

The intuitive idea behind the Theory of Types is as follows. We are given a collection of individuals. We further concern ourselves with classes of individuals, classes of classes of individuals, and so forth. Actually, we prefer at this point to talk of *properties* of individuals, properties of properties of individuals, and so forth. If we take an intensional view (speaking of properties rather than sets or classes) this will help to motivate our refusal to adopt an extensional criterion of identity between objects of positive type.

We begin to formalize this picture. We will develop TTU as a first-order theory with equality with sorts indexed by the natural numbers. The intuitive motivation is that type 0 is the type of individuals and type $n + 1$, for each concrete natural number n, is inhabited by properties of type n objects.

We supply countably many variables of each type. We do not encumber the variables with superscripts; we merely stipulate that each variable x has a unique type denoted by $\texttt{type}(x)$. This will make our notation more readable. Further, we provide a map `raise` on variables with the properties that $\texttt{type}(\texttt{raise}(x)) = \texttt{type}(x) + 1$ and that $\texttt{raise}(x) = \texttt{raise}(y)$ implies $x = y$. We may also denote

$\texttt{raise}(x)$ by x^+. Intuitively, raise represents the operation of incrementing a (suppressed) type superscript.

It is important to note that the use of numerals for the types is strictly a convenience. The types could equally well be presented as concrete tokens ι, ι', ι'' and so forth, and use of type superscripts would allow the map raise to be defined without any reference to arithmetic as well.

The only primitive non-logical predicate of TTU is membership, denoted by $\in$. Note that logical considerations tell us that an atomic formula $x = y$ is well-formed iff $\texttt{type}(x) = \texttt{type}(y)$. We stipulate further that an atomic formula $x \in y$ is well-formed iff $\texttt{type}(x) + 1 = \texttt{type}(y)$. This squares with our intuitive interpretatation of $x \in y$ as "x has property y"; if x is of type n, we expect y to be of type $n + 1$.

Note that we do not claim in TTU that objects of different types are different objects (we do not claim that the types are disjoint collections). We do not allow the question to be asked as to whether variables of different types represent the same object! Similarly, we do not allow ourselves to address the question of whether a type $n + 1$ property could be attributed to a type m object (where $m \neq n$).

We adopt an axiom scheme which asserts that any property of type n objects which we can express is actually realized by a type $n + 1$ object:

Axiom Scheme of Comprehension: For each formula ϕ, $(\exists A.(\forall x.x \in A \equiv \phi))$.

This is a scheme not only because of the infinite number of possible choices of formula ϕ, but also because of the infinite number of possible types for the variable x (the type of x determines the type of A).

Now we turn our attention to the issue of extensionality. It is possible to adopt strong extensionality in the Theory of Types (and Russell and Ramsey did this). We adopt an intensional viewpoint here which motivates us to be skeptical of extensionality; it may be that we will be accused of being disingenuous, since of course we know that the adoption of strong extensionality would bring us into collision with the unsolved problem of the consistency of *NF*. We will discuss this issue again below.

We denote the theory developed so far by TTU_0. We show how TTU_0 can interpret our target theory TTU, which has in addition a weak form of extensionality. This interpretation of extensionality is due to Marcel Crabbé in [3], where he used this approach to demonstrate that *NF* without any form of extensionality interprets *NFU*.

An intuitive approach to introducing extensionality is to point out that we can identify those type $n + 1$ objects which have the same elements. We define a relation to embody this idea:

Definition: We define $x \sim y$ for any x, y of the same positive type as $(\forall z.z \in x \equiv z \in y)$. For x, y of type 0, we allow $x \sim y$ to denote an unspecified equivalence relation (it may be equality, but we do not assume this).

We would like to use $\sim$ as our equality relation in each type. The difficulty which arises is that if $\sim$ is not the equality relation already, there will be properties in any type $n+2$ which do not respect the relation $\sim$ on type $n+1$. The solution to this problem which we adopt is to redefine the membership relation as well as the equality relation. We first introduce an auxilary predicate:

Definition: $\Sigma(x)$, read in English "x is a set", is defined as $(\forall yz.y \sim z \rightarrow (y \in x \equiv z \in x))$. This makes sense for x in any positive type (it doesn't make sense at type 0 because a type 0 variable cannot appear to the right of $\in$).

We say that x is a set just in case the property represented by x respects the identity relation $\sim$ of the appropriate type. This enables us to define a new membership relation:

Definition: $x \in_{\texttt{new}} y$ is defined as $x \in y \wedge \Sigma(y)$.

The new membership relation preserves the extensions of properties which respect the relations $\sim$ in each type, and assigns no elements at all to the properties whose extensions do not respect the equivalence relations $\sim$.

A property of any type which respects the relation $\sim$ (if $x \sim y$ then either both have the property or neither do) is assigned the same extension under $\in_{\texttt{new}}$ that it had under $\in$; a property of any type which does not respect the relation $\sim$ is assigned the empty extension. If we restrict ourselves to formulas using the predicates Σ, $\sim$, and $\in_{\texttt{new}}$, it is straightforward to verify that $\sim$ has the logical properties required to intepret the equality relation (substitution of "equals" for "equals" in any context preserves truth value).

The theory of Σ, $\sim$, and $\in_{\texttt{new}}$ in TTU_0 is an interpretation in TTU_0 of the inessentially stronger theory TTU which we now describe, in which $\sim$ interprets equality and $\in_{\texttt{new}}$ interprets membership.

TTU, like TTU_0, is a first order theory with equality and membership. It has the same sequence of sorts indexed by the natural numbers, and the same formation rules for atomic formulas, plus the primitive atomic formula $\Sigma(x)$ which is defined for x of any positive type. The axioms of TTU are

Sethood: $x \in y \rightarrow \Sigma(y)$

Extensionality: $(\Sigma(x) \wedge \Sigma(y)) \rightarrow (x = y \equiv (\forall z.z \in x \equiv z \in y))$

Comprehension: For each formula ϕ, $(\exists A.\Sigma(A) \wedge (\forall x.x \in A \equiv \phi))$.

Each of these "axioms" is actually a scheme because it comes in different versions for each appropriate type.

It is straightforward but tedious to verify that this theory is interpreted in TTU_0 in the way we have indicated. It is also easy to believe that this is the case, so we will not belabour the point!

An immediate advantage of TTU over TTU_0 is that we can define a canonical $\{x \mid \phi\}$ such that $\{x \mid \phi\}$ is a set and $(a \in \{x \mid \phi\}) \equiv \phi[a/x]$ (where $\phi[a/x]$ is the result of replacing x with a in ϕ). Comprehension ensures that there is at least one candidate for the role of $\{x \mid \phi\}$ for any formula ϕ; Extensionality ensures that there is at most one such candidate which is a set. All the non-sets have the same (empty) extension but there is a unique empty *set* $\emptyset = \{x \mid x \neq x\}$ in each type. In TTU_0 there is no way to choose a canonical object of a given extension (another way that one might do this, which would not require any extensionality assumptions, would be to adopt and apply the Axiom of Choice).

Of course, it is the case that we could have built a model of TT (the Theory of Types with strong extensionality $x = y \equiv (\forall z.z \in x \equiv z \in y)$). This process would have been logically more complex, because it would be necessary to eliminate the properties which did not respect the new equivalence relations, rather than simply make their extensions empty. Two *a priori* reasons to prefer the approach we have taken present themselves: the first is that this approach does not actually eliminate any objects from our model of "intensional type theory" TTU_0 (though it does collapse some distinctions between them); the second is that the construction for a model of TT is intrinsically considerably more complicated. It could not be presented (as this one has) as a scheme of definitions of the same form at each type; the definition of the interpreted TT would in fact be a recursively defined scheme with the definition of the new relations at each type depending on the new relations on the preceding type. (It is worth noting that reasoning in the theory of types involving such recursion on the type structure *cannot* be reproduced in the system *NFU* we will develop.)

In any event, we temporarily adopt TTU as our foundation for mathematics. In the next section, we will discuss the actual development of some mathematics in TTU (and the mathematical development will motivativate us to adopt two additional axioms, the axioms of Infinity and Choice).

We review the intuitive picture behind TTU. We have type 0 (our featureless "individuals") to start with; each type $n+1$ consists of sets or classes of type n objects plus possibly some additional objects (which we will call *urelements* or *atoms*; these are not to be confused with the individuals of type 0).

3 The Development of Mathematics in TTU

In this section, we will develop some basic mathematics in TTU. The development will go in the same way as it would in TT (our weakening of extensionality does not affect the suitability of the system for foundations). It will also become clear in the course of this development why the type structure of TTU (or TT) can be regarded as annoying. This will help to motivate our eventual revision of the theory to obtain type-free foundations in *NFU*. The mathematical development will also motivate the addition of the axioms of Infinity and Choice.

We begin by defining the natural numbers. The intuitive idea is that we will define each concrete natural number n as the set of all sets with n elements. This must immediately be modified by type considerations: we will define n (of type 2) as the set of all type 1 sets (of type 0 individuals) with n elements. Then, of course, we need to verify that we can actually do this.

We define 0 as $\{\emptyset\}$, the set whose only element is the (type 1) empty set. Recall from above that there may be many elements of type 1 with no elements (because of the weakening of extensionality in TTU), but there will be exactly one *set* with no elements.

For any (type 1) set A, we define $A + 1$ as $\{a \cup \{x\} \mid a \in A \wedge x \notin a\}$; $A+1$ is defined as the set of all disjoint unions of elements of A with singletons. Observe that 0+1 will be the set 1 of all singletons, 1+1 will be the set 2 of all sets with two elements, 2+1 will be the set 3 of all sets with 3 elements, and so on through the concrete natural numbers.

We define $\mathcal{N}$ as $\{n \mid (\forall \mathcal{A}.(0 \in \mathcal{A} \wedge (\forall A.A \in \mathcal{A} \rightarrow A + 1 \in \mathcal{A})) \rightarrow n \in \mathcal{A})\}$. $\mathcal{N}$ is the (type 3) set of all (type 2) sets which belong to all (type 3) sets which contain 0 and are closed under our "successor" operation. This is a reasonable definition of the set of natural numbers.

Notice that this is an impredicative instance of comprehension: if one thinks of the instance of comprehension providing us with $\mathcal{N}$ as a definition of $\mathcal{N}$, it is disturbing that $\mathcal{N}$ itself falls within the scope of the quantifier over $\mathcal{A}$ ($\mathcal{N}$ is "defined" as the intersection of a collection of sets to which it itself belongs). We admit to no philosophical qualms about this (we do not think that instances of comprehension are definitions); there is a little more discussion of impredicativity later in the paper.

An annoying feature of this definition is that it must be repeated in exactly the same way if one wishes to "count" objects of types higher than 0. There is a type 4 set $\mathcal{N}$ which provides us with numerals for counting type 2 objects, and so forth in each higher type. It is not usual to have to have different numerals to count objects of different sorts, though this is true in some natural languages (e.g. Japanese).

We have not assumed an axiom of infinity at this point. If there were finitely many type 0 objects, we would discover that the type 1 set $V = \{x \mid x = x\}$ would belong to some natural number. An axiom of infinity excluding this situation could look like this:

***Axiom of Infinity:** $(\forall n \in \mathcal{N}.V \notin n\})$

It is straightforward to prove that the version of this axiom with V at type 1 implies the analogous assertions at each higher type. The star indicates that we do not actually adopt this as an axiom; it is a consequence of the Axiom of Ordered Pairs introduced below.

We now consider the definition of the ordered pair. The following standard definition (due to Kuratowski) could be used:

***Definition:** $\langle x, y\rangle$, read "the ordered pair of x and y" is defined as $\{\{x\}, \{x, y\}\}$.

This definition (which we have starred to indicate that we do not ultimately adopt it) has the technical disadvantage that the pair is two types higher than its projections. This has the odd effect that functions are three types higher than their values and arguments. We prefer to have a pair which is of the same type as its projections. If we were using TT (i.e., if we assumed strong extensionality) we would be able to define such a pair on all sufficiently high types (this is due to Quine in [23]). In TTU it is not possible to define the type-level pair, but it is possible to assume that there is one as long as the Axiom of Infinity is assumed. The precise situation is as follows: if we assume the Axiom of Choice (as we will), it is possible to prove that there is a type level pair on each type (this is a consequence of the theorem $\kappa^2 = \kappa$ of transfinite cardinal arithmetic); if we did not assume choice, it would still be possible to demonstrate that TTU with Infinity interprets TTU with a type-level pair (this development would exploit Quine's definition of the pair for pure sets). Since we do assume choice, we see no reason to go into details. We avoid the necessity of developing the theory of relations and functions with the Kuratowski pair far enough to prove $\kappa^2 = \kappa$ by adopting the type-level pair as a primitive of our theory.

Here are the formal details:

We add predicates π_1 and π_2 to our formal language. $x \, \pi_1 \, y$ and $x \, \pi_2 \, y$ are well-formed iff $\texttt{type}(x) = \texttt{type}(y)$, and satisfy the following

Axiom of Ordered Pairs: For each x, y (of any type) there is a unique object z (of the same type) such that $z \, \pi_1 \, x$ and $z \, \pi_2 \, y$. We will denote this uniquely determined object by $\langle x, y \rangle$.

It is easy to show that the Axiom of Ordered Pairs implies the Axiom of Infinity.

Once the ordered pair is introduced, we can develop the theory of functions and relations in the usual way. An advantage of the type-level pair is that relations are one type higher than the elements of their domains and ranges, and functions are one type higher than their arguments and values; if we used the Kuratowski pair, these displacements would be equal to 3. The Axiom of Choice can be introduced in a number of ways. A common form of the Axiom of Choice looks like this:

***Axiom of Choice:** Let P be a collection of pairwise disjoint nonempty sets. Then there is a set C (of type one lower than the type of P) which contains exactly one element of each element of P.

The usual equivalences between forms of the Axiom of Choice are provable, including the equivalence with the assertion that all sets can be well-ordered. We prefer to use the latter form, and in the context of type theory it is sufficient to say that each type can be well-ordered. So we introduce a relation symbol $\leq$ and stipulate that $x \leq y$ is a well-formed atomic formula iff $\texttt{type}(x) = \texttt{type}(y)$. Our official axiom is:

Axiom of Choice: $\leq$ is a well-ordering (of each type). (That is, $\leq$ is a linear order, and any nonempty set has a $\leq$-least element).

Two sets A and B are said to be equinumerous just in case there is a bijection between them (just as in the usual set theory). It is natural in type theory to define the cardinal number $|A|$ as the set of all sets equinumerous with A. Notice that $|A|$ is one type higher than A, and also that the finite cardinal numbers will coincide with the natural numbers already defined. Ordinal numbers are usually defined in type theory as equivalence classes of well-orderings under similarity.

We briefly review the paradoxes of naive set theory. The Russell paradox of the class $\{x \mid x \notin x\}$ cannot be reproduced in type theory because the formula $x \notin x$ cannot be well-formed (no matter what the type of x). The Cantor paradox of naive set theory applies the theorem $|A| < |\mathcal{P}(A)|$ (the cardinality of a set is strictly smaller than the cardinality of its power set) to the cardinality of the universe to obtain an absurd result; this doesn't work because $|A| < |\mathcal{P}(A)|$ is ill-typed: the set A and the set $\mathcal{P}(A)$ of all subsets of A are at different types. To fix this, we introduce a

Definition: We define $\mathcal{P}_1(A)$ as the set of all one-element subsets of A.

We *can* prove $|\mathcal{P}_1(A)| < |\mathcal{P}(A)|$ in TTU, from which we can prove $|\mathcal{P}_1(V)| < |\mathcal{P}(V)|$: the set of all one-element subsets of the universal set V of all elements of a particular type is smaller than the set of all subsets of V (both of these sets live in the next type above the type of V). The Burali-Forti pardox of the largest ordinal is again avoided by type considerations. The set of all ordinal numbers in a given type is well-ordered by the usual order on ordinal numbers, and so belongs to an ordinal number, but this ordinal is of a higher type than any of the ordinals in the collection of ordinals we started with.

Hereinafter we take TTU to include the axioms of Ordered Pairs and of Choice.

4 Model Theory of *TTU* in *TTU*

We will use TTU as our vehicle for metamathematics as well as for mathematics. It should be clear that TTU is adequate to prove standard results in model theory. In this section, we discuss the definition of the notion "model of TTU" inside TTU. Of course, it will not be possible to prove inside TTU that there is a model of TTU (any more than it is possible to prove inside *ZFC* that there is a model of *ZFC*!).

A model of TTU is determined by a sequence of sets T_i implementing the types plus relations implementing membership between each pair of successive types, the projection relations on each type, and the well-orderings of each type. If we assumed that the T_i's were pairwise disjoint, we could represent the membership relation of the model by a single relation E, but we prefer not to make this assumption: thus, we provide a sequence of membership relations $E_i \subseteq T_i \times T_{i+1}$ coding membership of type i objects in type $i+1$ objects in the object theory. Notice that the "types" T_i are all sets of the same type $n+1$ in the metatheory: the relations E_i are at the same type as the T_i's, the elements of any type of the model are one type lower, and the sequences T and E are one type higher. We will implement a model as a structure $\langle T, E, \mathtt{P1}, \mathtt{P2}, W \rangle$, where T is the sequence of types, E is the sequence of membership relations, P1 and P2 are sequences of first and second projection relations for types of the model, and W is a sequence of well-orderings of types of the model. We may abuse notation by referring to a model as $\langle T, E \rangle$ when no reference to the other relations is needed.

We do not formally define satisfaction of sentences by the model (standard methods are readily adapted to type theory). Of course, there are additional conditions in the definition of the notion "model of TTU" which express the conditions that it satisfy each axiom of TTU! The finite axiomatization of NFU given below suggests a simple way to define "model of TTU" which would not even require a formal definition of satisfaction of sentences; it is straightforward to express satisfaction of the typed version of each of those axioms by a model without metamathematical finesse.

We do define a particular class of models of special interest: we say that a model $\langle T, E \rangle$ of TTU is *natural* just in case for each natural number i and subset A of T_i there is an element a of T_{i+1} such that for all $x \in T_i$, $x \in A$ if and only if $x \, E_i \, a$; in a natural model, every subset of each type of the model is coded by some element of the next higher type of the model.

As we remarked above, the incompleteness theorems show us that we cannot hope to prove the existence of a model of TTU (natural or otherwise) in TTU. However, we do know (by standard results of model theory) that if TTU is consistent, there will be a countable (and so of course not natural) model of TTU. The assumption that there is a natural model of TTU is presumably consistent with TTU, but strengthens the theory (in much the same way that the assumption that there is an inaccessible cardinal strengthens ZFC, though we are talking here about weaker theories).

If we define TTU_n, for each concrete natural number n, as the fragment of TTU obtained by restricting our attention to types 0-$(n-1)$ of TTU, it turns out that we can prove the existence of natural models of TTU_n for each concrete n, though of course not in a uniform manner. We sketch the construction of such a natural model of TTU_n. The key idea, from which all the details of the model can be worked out, is that for any types i and j with $i < j$, there is a natural way to code type i objects into type j: an object x of type i can be coded by its $(j-i)$-fold iterated singleton. The natural model of TTU_n will have as the elements of its types 0-$(n-1)$ the elements of the true types 0-$(n-1)$ as coded into type $n-1$. Thus, the set T_i for each i will be the type n set of all $((n-1)-i)$-fold singletons of objects of type i. The "membership relation" E_{n-2} between the two highest types T_{n-2} and T_{n-1} will be the intersection of the subset relation with $T_{n-2} \times T_{n-1}$: if we have x of type $n-2$ and y of type $n-1$, x is coded by $\{x\}$ and y is coded by itself, and $x \in y$ iff $\{x\} \subseteq y$. In general, the relation E_i for $0 \leq i \leq n-2$ will be the type n relation on iterated singletons in type $n-1$ induced by the intersection of the subset relation with $T_i \times T_{i+1}$. It is easy to see that this is a natural model, because collections of iterated singletons of type i objects correspond precisely to collections of type i objects.

5 Typical Ambiguity in *TTU*

The Theory of Types was proposed before the development of Zermelo-Fraenkel set theory, but it was not adopted generally as a foundation for mathematics. The reasons for this have to at least partially with the cumbersome nature of a notation cluttered with type superscripts; in our development we have avoided this problem. A further reason, which we have not been able to avoid, is the high level of polymorphism in this theory.

Recall that when we defined the set $\mathcal{N}$, we realized that we had succeeded in defining numerals for counting type 0 objects, but that we would need to define a different set $\mathcal{N}$ in a different type for the counting of objects of each type. The natural number 3 (in type 2) is the set of all three-element sets of type 0 objects, but there are further natural numbers 3 which are sets of all three element sets of type 1 objects, sets of all three element sets of type 186 objects, etc. Actually, we can't really say that these are "further" numbers 3; the syntax of our language forbids us from comparing any two of these, or from considering the sequence of "numbers 3" except on the meta-level (it is easy to define the sequence of "numbers 3" in a model of TTU).

This is a particular case of a general phenomenon, known to Russell, called "typical ambiguity". We introduce a useful

Definition: For any formula ϕ of the language of TTU, define ϕ^+ as the formula which results from the replacement of each variable x occurring in ϕ with $\texttt{raise}(x)$.

It should be evident that ϕ^+ will itself be a well-formed formula. This notation allows us to describe the phenomena of typical ambiguity succinctly: for each object $\{x \mid \phi\}$ that we can define, there is a precisely analogous object $\{\texttt{raise}(x) \mid \phi^+\}$ which we can define in the next higher type. The ambiguity is even more profound. It is easy to see that for each axiom ϕ (including instances of Ordered Pairs and of Choice), ϕ^+ is also an axiom: from this it follows easily that for every theorem ϕ, ϕ^+ is also a theorem.

The type structure of TTU looks rather like a hall of mirrors! The reduplication of objects and theorems in each type suggests that perhaps what is going on is that the types (which we have never assumed to be distinct) are actually all identical. This is the motivation for Quine's original proposal (based in TT rather than TTU) of the theory "New Foundations". The difficulty with Quine's suggestion (which still has not been justified in its original form) is that there is no obvious reason to believe that this can be the case. In fact, if we think of the "natural" models of TT or TTU in Zermelo-style set theory this seems absurd: type $n+1$ in a natural model will be (or include if there are urelements in the type) the power set of type n, and so cannot be the same as type n (because it must be strictly larger than type n by Cantor's theorem). What happens in natural models inside TTU is technically a bit different, but the outcome is the same: a natural model of TTU in TTU will have types all of different cardinalities and so distinct.

Our answer to this objection will become plainer later, but for the moment we merely say that we have no particular reason to believe that our "ultimate" model of TTU is a natural model (in the sense that type $n+1$ contains every arbitrary subcollection of type n). The intensional view of type theory that we suggested when motivating the weakening of extensionality also leaves us open to skepticism as to whether every arbitrary subcollection of type n is realized by a property of type n objects (an element of type $n+1$). Certainly if we want to pursue the possibility of a hidden identity between the types, we must abandon the idea that type $n+1$ contains an object representing each arbitrary subcollection of type n.

In any event, mathematical convenience strongly suggests that we would wish to collapse the type structure of TTU.

6 A Road Not Taken: Collapsing *TTU* to Mac Lane Set Theory with Atoms

It should be noted before we proceed to our official technique of collapsing the type structure that there is a natural way to collapse the type structure of TTU which leads to a Zermelo-style set theory (though not all the way to ZFC). This collapse works best in models of TT; there can be a technical obstruction to collapsing models of TTU with urelements, which is removed if we stipulate for each type $n+1$ that there are at least as many urelements of type $n+1$ as there are singletons of type n urelements (i.e., as we go up in type the number of urelements added at each type will not decrease). This given, we proceed as follows: we are initially given an injection (one-to-one map) f_0 from the singletons of type 0 objects to type 1 objects (the role of the singleton operation here is just to make it possible to type the map f_0 within TTU); for each natural number n, once we have defined f_n, an injection from singletons of type n objects into type $n+1$ objects, we define f_{n+1} as follows: for any set A in type $n+1$, we define $f_{n+1}(\{A\})$ as $\{f_n(\{a\}) \mid a \in A\}$ (the elementwise image of $\mathcal{P}_1(A)$ under f_n), and choose an arbitrary injection from singletons of type $n+1$ urelements to type $n+2$ urelements to serve as the restriction of f_{n+1} to urelements. This system of maps f_n can be used to interpret an untyped set theory in TTU: each f_n is used in the obvious way to identify the type n objects with a subset of the type $n+1$ objects, and the identification maps are defined in such a way as to respect membership. The untyped set theory which results is Zermelo set theory with atoms, with the restriction that the axiom of separation only applies to formulas in which each quantifier is restricted to a set; this theory can also be called "Mac Lane set theory with atoms". (For Mac Lane's proposal of this theory, see [19]; for an excellent recent study of this theory, see [20]). The restriction on separation arises essentially because TTU never allows us to quantify over all types at once. The model of Mac Lane set theory obtained in this way might turn out to be a model of Zermelo set theory, but cannot turn out to be a model of ZFC, because the sequence of sets implementing types of the original model will not be a set and will provide a counterexample to replacement (if we get a model of ZFC, that implies that we had no urelements to start with, and under this condition the sequence of types of the original model is definable in the model of Mac Lane set theory; it may actually be possible to get a model of ZFA (ZFC weakened to allow atoms) by collapsing a model of TTU, because it is not clear that the sequence of types is definable in the general situation; in any case it is possible that the model obtained by collapsing a general model of TTU (even an extensional one) will contain an inner model of ZFC, if it happens to contain an inaccessible cardinal).

The technical details of the collapse described in the preceding paragraph are not important to our development, but the fact that it is possible to get to Zermelo-style foundations by collapsing types in TTU (more naturally from TT) should be reviewed in this context for comparison and contrast with the rather different collapse of types which leads to NFU. The logical complexity of the collapse sketched in the previous paragraph is about the same as the complexity of the interpretation of type theory with full extensionality in TTU_0 which we briefly sketched earlier; it clearly involves recursion on types.

7 Polymorphism Used to Collapse the Type Structure

If we wish to collapse the type structure in such a way as to exploit the polymorphism of TTU, we obtain quite different results from those of the construction of the previous section. It is not generally the case that the collapse defined in the previous section will identify a set $\{x \mid \phi\}$ with its analogue $\{x^+ \mid \phi^+\}$ in the next higher type. This will be true for finite well-founded sets, but not generally for any others. For example the "numbers 3" of a model of TTU will be collapsed to the sets $[V_i]^3$ in the model of Mac Lane set theory with atoms, where the set V_i is the set coding type i and $[X]^3$ represents the set of all three-element subsets of X. The sets $[V_i]^3$ are demonstrably distinct, and this is a typical situation.

To assume that such a collapse is possible has nontrivial logical content, expressible as an axiom scheme to be adjoined to TTU. To see this, consider the case of set abstracts $\{x \mid \phi\}$ where the formula ϕ happens not to contain any free occurrences of x, or any parameters (i.e., ϕ is a sentence). Any such set $\{x \mid \phi\}$ will be equal to the universe or the empty set depending on whether ϕ happens to be true or false. $\{x^+ \mid \phi^+\}$ will have the same property, and we certainly expect the universe in a type to be analogous to the universe in the next type, and similarly for the empty set. It follows that we must believe the following axiom scheme in TTU:

Ambiguity Scheme: For any sentence ϕ, $\phi \equiv \phi^+$

The truth of this scheme is clearly necessary for a polymorphic collapse to succeed; it has been shown that it is in a sense also sufficient (see [27] and a later section of this paper (but we cheat slightly in our proof)).

We now define the theory *NFU* which results from the identification of the types in TTU. We should point out as we do this that we have not yet justified this maneuver; we can see what theory results from collapsing the type structure of TTU, but we cannot yet see intuitively or otherwise that this theory is legitimate.

NFU is a first-order theory with equality, membership, the sethood predicate Σ, projection relations π_1 and π_2 and a well-ordering $\leq$. Its axioms are Sethood, Extensionality (the weak form given above for TTU), Comprehension, Ordered Pairs, and Choice, as in TTU but with all distinctions of type ignored. The only one of these axioms which requires comment is Comprehension.

It may seem that the form of Comprehension which we obtain by dropping all indications of type from the Comprehension scheme of TTU is

***Axiom Scheme of Comprehension:** For each formula ϕ (of the language of *NFU*) $(\exists A.\Sigma(A) \wedge (\forall x.x \in A \equiv \phi))$.

Of course this scheme is false: consideration of the Russell class $\{x \mid x \notin x\}$ is sufficient to see this. But there is a fallacious assumption here (explicitly supplied as a parenthesis). When we omit types from the Comprehension scheme of TTU, we do not obtain an axiom $(\exists A.\Sigma(a) \wedge (\forall x.x \in A \equiv \phi))$ for each formula ϕ of the language of *NFU*: we obtain an axiom for each formula ϕ of the language of *NFU* which can be obtained by ignoring distinctions of type in a formula of the language of TTU. The formula $x \notin x$, for example, cannot be obtained in this way.

It does not seem satisfactory to define an axiom in an untyped theory by essential reference to a typed theory, so we express this a little differently with the help of a

Definition: We say that a formula ϕ in the language of *NFU* is *stratified* just in case there is a function `type` (which we call a *stratification* of ϕ) from the variables of the language of *NFU* to the natural numbers with the property that for each atomic subformula $x = y$ or $x \, \pi_i \, y$ of ϕ we have $\texttt{type}(x) = \texttt{type}(y)$ and for each atomic subformula $x \in y$ of ϕ we have $\texttt{type}(x) + 1 = \texttt{type}(y)$.

Any similarity between this definition and the well-formedness conditions for formulas of TTU is far from accidental. A stratified formula ϕ will be obtainable from a formula of the language of TTU by disregarding type distinctions; we can now present the

Axiom Scheme of Stratified Comprehension: For each stratified formula ϕ of the language of *NFU*, $(\exists A.\Sigma(a) \wedge (\forall x.x \in A \equiv \phi))$.

This is still a complex axiom scheme, and of course we still see the genetic relationship of *NFU* to type theory, but at least we have avoided the disaster of unstratified comprehension. But we are certainly open at this point to the accusation that we have perpetrated a "syntactical trick"!

(We note at this point that we ourselves question whether even Quine was guilty of a mere "syntactical trick". We would describe Quine as guilty of being unduly hopeful that the identification of the types suggested by the typical ambiguity of *TT* represented a real phenomenon.)

8 *NFU* is Finitely Axiomatizable

In this section we demonstrate that the Stratified Comprehension scheme of *NFU* can be replaced with finitely many of its instances in a natural way; i.e., *NFU* is finitely axiomatizable, and in a way which makes no reference to types at all. The finite axiomatization of stratified comprehension [9] which is usually referenced is rather difficult to understand; the one we present here (taken from our [13], with a considerable simplification of the result analogous to Lemma 1) is much more natural. The simplification is largely due to the use of a type-level pair (though not entirely; it is fairly easy to modify this axiom set to use the Kuratowski pair, and the result is still more appealing than the axiom set of [9]).

Here is the set of axioms. Recall that $\{x \mid \phi\}$ is defined as the unique object A (if there is one) such that $x \in A \equiv \phi$. We also use the notation $\{\langle x, y\rangle \mid \phi\}$ and similar notations with the usual meaning.

Axiom of Sethood: As above.

Axiom of Extensionality: As above.

Axiom of the Universal Set: $V = \{x \mid x = x\}$ exists.

Axiom of Complements: $A^c = \{x \mid x \notin A\}$ exists.

Axiom of Boolean Unions: $A \cup B = \{x \mid x \in A \vee x \in B\}$ exists.

Axiom of Set Union: $\bigcup A = \{x \mid (\exists y. x \in y \wedge y \in A)\}$ exists.

Axiom of Singletons: $\{A\} = \{x \mid x = A\}$ exists.

Axiom of Ordered Pairs: As above.

Axiom of Cartesian Products: $A \times B = \{\langle x, y\rangle \mid x \in A \wedge y \in B\}$ exists.

Axiom of Converses: $R^{-1} = \{\langle x, y\rangle \mid \langle y, x\rangle \in R\}$ exists.

Axiom of Relative Products: $R|S = \{\langle x, y\rangle \mid (\exists z. \langle x, z\rangle \in R \wedge \langle z, y\rangle \in S)\}$ exists.

Axiom of Domains: $\text{dom}(R) = \{x \mid (\exists y. \langle x, y\rangle \in R)\}$ exists.

Axiom of Singleton Images: $R^\iota = \{\langle \{x\}, \{y\}\rangle \mid \langle x, y\rangle \in R\}$ exists.

Axiom of the Diagonal: The set Eq representing the equality relation, $\{\langle x, x\rangle \mid x = x\}$, exists.

Axiom of Projections: The sets P1 and P2 representing the projection relations, $\{\langle\langle x, y\rangle, x\rangle \mid x = x\}$ and $\{\langle\langle x, y\rangle, y\rangle \mid x = x\}$, exist.

Axiom of Inclusion: The set Subset representing the subset relation, $\{\langle x, y\rangle \mid \Sigma(x) \wedge \Sigma(y) \wedge (\forall z. z \in x \rightarrow z \in y)\}$, exists.

Axiom of Choice: As above.

The first thing to notice about this axiom set is that the sets and operations on sets provided by the axioms are all intuitively reasonable (even the universal set and complements, if one does not have prior training in *ZFC*): they are the primitives of boolean and relation algebra, plus the domain operator needed to reduce binary relations to sets, the projection and subset relations, and the special set operations of singleton, singleton image, and set union. The axiom of the universal set is actually redundant; one can use any specific set introduced in the other axioms along with the boolean operations to construct V.

Our aim for the rest of this section is to prove the following

Meta-Theorem: For any stratified formula ϕ, $\{x \mid \phi\}$ exists.

Lemma 1: Suppose that ϕ is built up by logical operations from sentences of the forms $u \in A$ and $\langle u, v\rangle \in R$ where A and R are parameters or constants. Then $\{x \mid \phi\}$ and $\{\langle x, y\rangle \mid \phi\}$ exist.

Proof of Lemma 1: We prove this by induction on the structure of ϕ. If ϕ is of the form $\neg\psi$, then we construct $\{x \mid \phi\}$ as the complement of $\{x \mid \psi\}$, and similarly for $\{\langle x, y\rangle \mid \phi\}$. If ϕ is of the form $\psi \vee \chi$ we construct $\{x \mid \phi\}$ as the boolean union of $\{x \mid \psi\}$ and $\{x \mid \chi\}$, and we prove the existence of $\{\langle x, y\rangle \mid \phi\}$ in the same way. If ϕ is of the form $(\exists z.\psi)$, we construct $\{x \mid \phi\}$ as the domain of $\{\langle x, z\rangle \mid \psi\}$.

The case that remains is the hard one. We construct $\{\langle x, y\rangle \mid (\exists z.\psi)\}$. In order to do this, we expand our language to include terms obtained by possibly repeated application of the projection functions π_1 and π_2 to variables. We will indicate below how these are eliminated when we get to the stage of atomic formulas. Notice that any composition of projection operators can be constructed as a relation using the axiom of relative products. Once this expansion of our language is made, we define the set $\{\langle x, y\rangle \mid (\exists z.\psi)\}$ as the domain of $\{\langle u, z\rangle \mid \psi'\}$, where u is a new variable and ψ' is obtained from ψ by replacing each occurrence of x with the term $\pi_1(u)$ and each occurrence of y with the term $\pi_2(u)$.

We may suppose without loss of generality that no other logical connective or quantifier occurs in ϕ.

At the stage of atomic formulas, we need to consider first "atomic formulas" of the forms $\pi_a(x) \in B$ and $\langle\pi_a(x), \pi_b(y)\rangle \in C$, where the operators

π_a and π_b are compositions of projection operators (`Eq` may be used as the composition of zero projection operators (the identity map) if needed; also the variables x and y are not necessarily distinct). First note that any composition of projection operators is realized by a set relation by application of the axioms of projections and relative products. Thus, we can convert $\langle \pi_a(x), \pi_b(y) \rangle \in C$ to the form $\langle x, y \rangle \in (\pi_a|C|\pi_b^{-1})$ and $\pi_a(x) \in B$ to the form $x \in \mathrm{dom}(\pi_a \cap (V \times B))$, producing atomic formulas without the new function symbols.

All that remains is to verify the theorem for atomic formulas of the forms listed above. In what follows, variables are distinct unless written with the same letter. $\{u \mid u \in A\}$ is equal to A if A is a set and to the empty set otherwise. $\{\langle u, v \rangle \mid u \in A\}$ is $A \times V$; $\{\langle v, u \rangle \mid u \in A\}$ is $V \times A$. $\{v \mid u \in A\}$ and $\{\langle v, w \rangle \mid u \in A\}$ are either the universe (resp. universal relation) or the empty set, depending on whether $u \in A$.

$\{u \mid \langle u, v \rangle \in R\}$ is the domain of the intersection of R and $V \times \{v\}$ (we have intersections because we have unions and complements). $\{v \mid \langle u, v \rangle \in R\}$ is the domain of the converse (the range) of the intersection of R and $\{u\} \times V$. $\{w \mid \langle u, v \rangle \in R\}$ is the universe or the empty set. $\{u \mid \langle u, u \rangle \in R\}$ is the domain of the intersection of R and `Eq`. $\{v \mid \langle u, u \rangle \in R\}$ is the universe or the empty set. $\{\langle u, v \rangle \mid \langle u, v \rangle \in R\}$ is the intersection of R and $V \times V$. $\{\langle v, u \rangle \mid \langle u, v \rangle \in R\}$ is the converse of the intersection of R and $V \times V$. $\{\langle u, w \rangle \mid \langle u, v \rangle \in R\}$ is the cartesian product of $\{u \mid \langle u, v \rangle \in R\}$ (already shown to exist) and V, and similar considerations apply to the three other cases in which the two pairs share one variable. If the two pairs share no variable, we have the universal relation or the empty set. $\{\langle u, v \rangle \mid \langle u, u \rangle \in R\}$ and $\{\langle v, u \rangle \mid \langle u, u \rangle \in R\}$ are cartesian products with V of $\{u \mid \langle u, u \rangle \in R\}$, which has already been shown to exist.

The proof of Lemma 1 is complete. For the idea of exploiting the projection operators to make it possible to consider only sets and binary relations, we are indebted to Tarski and Givant ([28]).

Lemma 2: The set $E = \{\langle \{x\}, y \rangle \mid x \in y\}$ exists.

Proof of Lemma 2: The domain of $(V \times V)^\iota$ is the set of all singletons, which we will call 1. $(1 \times V) \cap \texttt{Subset} = E$. It is useful to note that this relation played an important role in the construction of natural models of TTU_n above.

Definition: Define V^0 as V; define V^{n+1} as the domain of $(V^n \times V^n)^\iota$. V^n will be the set of all n-fold iterated singletons. (This is not to be confused with the standard notation V_α for stage α of the cumulative hierarchy in ZFC, which will be used below). For any relation R taken from the list (`Eq`, `P1`, `P2`, E), define R_0 as R and R_{n+1} as R_n^ι for each concrete n; R_n is the relation on n-fold iterated singletons induced by R.

Proof of Meta-Theorem: Let ϕ be an arbitrary stratified formula, and let `type` be a stratification of ϕ with an upper bound N on its range. We transform ϕ into a related formula ϕ' which will have the form required by Lemma 1. ϕ' is constructed by replacing each occurrence of $x \in y$ with $\langle x, y\rangle \in E_{N-\mathtt{type}(y)}$, replacing each occurrence of $x = y$ with $\langle x, y\rangle \in \mathtt{Eq}_{N-\mathtt{type}(x)}$, where $\mathtt{Eq}$ is the set provided by the axiom of equality, and replacing occurrences of $x\,\pi_i\,y$ with $\langle x, y\rangle \in \mathtt{Pi}_{N-\mathtt{type}(x)}$ (where $\mathtt{Pi}$ stands for $\mathtt{P1}$ or $\mathtt{P2}$), and restricting each quantifier over a variable x to the set $V^{N-\mathtt{type}(x)}$. Occurrences of $\Sigma(x)$ are best eliminated by replacement with $(\exists y.y \in x) \vee x \in \{\emptyset\}$, then carrying out the translation of this as above.

The motivation behind this transformation is that each variable x in ϕ has different reference in ϕ': x refers in ϕ' to the $(N - \mathtt{type}(x))$-fold singleton of its referent in ϕ. It is useful to note that this is essentially the same as the coding trick used in the construction of natural models of TTU_n above: the relations used to code membership at different (relative) types are the same as in that construction. In any case, it is easy to check that the changes in relations and the restrictions of quantifiers are those which should be induced by the intended change of reference of variables, if the truth values of ϕ and ϕ' are to be the same.

For any variable x, $\{x \mid \phi'\}$ exists by Lemma 1. This will be the set of $(N-\mathtt{type}(x))$-fold singletons of objects x such that ϕ; apply the operation of set union $(N - \mathtt{type}(x))$ times to this set to get $\{x \mid \phi\}$.

The proof of the Meta-Theorem is complete.

The presentation of *NFU* in this section has the advantage that there is no metamathematics in the formulation of the comprehension axioms; each axiom tells us of the existence of concrete sets or operations on sets. It increases the sense that stratified comprehension may be a reasonable criterion for set existence, but it certainly does not provide firm support for this impression. After all, it is known that if we add strong extensionality (another intuitively appealing assumption) to the given axiom set it becomes inconsistent (because we would get *NF* + Choice, which was shown to be inconsistent by Specker in [26]).

Another practical reason to be interested in this section is that it suggests a clean way to define the notion "model of *TTU*"; the axiomatization given here, if typed, gives an axiomatization of *TTU* as well, and it is much easier to describe satisfaction of these axioms in a model of *TTU* in our metatheory than to describe satisfaction of the general comprehension scheme. We don't carry out the details here, but it is useful to be aware of the issue.

Further, the details of this proof are important in developing and understanding the intuitive picture of set theory we develop below.

9 Sufficiency of the Ambiguity Scheme

In this section, we prove an important theorem of Specker.

Theorem (Specker): If *TTU* + the Ambiguity Scheme is consistent, then *NFU* is consistent.

Proof of Theorem: We use the well-orderings provided by the Axiom of Choice to construct a term model of TTU. Note that any sentence $(\exists x.\phi)$ which is true has a witness $(\mu x.\phi)$ definable as the $\leq$-least element of $\{x \mid \phi\}$. When $(\exists x.\phi)$ is not true, we might as well define $(\mu x.\phi)$ as $\emptyset$. Take any model of the revised TTU: it is straightforward to show that the substructure consisting of the referents of all μ-terms is a model of TTU as well. (Another way of putting all this is that we are exploiting the fact that TTU has definable Skolem functions.)

A model of *NFU* is then obtained by identifying the analogous μ-terms of different types in the term model constructed in this way from a model of TTU + Ambiguity. There can be no conflict between statements true of analogous μ-terms at different types of TTU, by Ambiguity (the crucial point here is that all elements of the term model are definable!). The semantics of stratified sentences of *NFU* are obtained directly from the term model (or indeed from the original model of TTU + Ambiguity, which has the same theory), and the semantics of unstratified sentences are obtained from the term model in the natural way.

The proof is complete.

The proof of the sufficiency of ambiguity is much easier in TTU as we have presented it, because the presence of Choice gives us definable witnesses to every true existential statement (even with parameters – thus definable Skolem functions). A similar technique will work for theories without Choice, but may appear to be cheating: it adds no strength to type theory without Choice to add a well-ordering of each type but exclude the well-ordering from instances of comprehension (the well-orderings can be interpreted as external well-orderings of each type in a countable model). If one extends the Ambiguity Scheme to formulas containing the external well-ordering, the argument goes just as it does here. It turns out that one does not need to cheat: the equiconsistency of type theories with ambiguity schemes and the corresponding *NF*-like systems is provable anyway, in the absence of definable Skolem functions (using saturated models); the curious reader can see the details in [27] or [5], p. 59.

10 Consistency of *NFU* from Ramsey's Theorem

In this section, we will give what is essentially Jensen's proof that *NFU* is consistent, but in a form proposed by Maurice Boffa ([1]). We feel that this proof is not especially intuitively appealing, but it certainly does work.

It uses Ramsey's theorem, which we now state:

Definition: For any set X, we define $[X]^n$ as the set of all n-element subsets of n. Let P be a finite set of pairwise disjoint sets whose union is $[X]^n$ (a finite partition of $[X]^n$). We say that $H \subseteq X$ is a *homogeneous set for P* iff there is $p \in P$ such that $[H]^n \subseteq p$; in other words, all n-element subsets of H fall in the same "compartment" of the partition P.

Theorem (Ramsey): For any infinite set X and partition P of $[X]^n$, there is an infinite homogeneous set for P.

This theorem can be proved in TTU (in various typed versions) in essentially the same way that it is proved in standard set theory.

The main result of this section is

Theorem: *NFU* is consistent iff TTU is consistent.

It is obvious that TTU is consistent if *NFU* is consistent; given a model of *NFU* with domain M and membership relation e (in some type in TTU, which we recall is our metatheory), we can define a model in the following way: $T_i = M$ for each i and $E_i = e$ for each i, and the sequences of projection operators and well-orderings will be constant sequences of the projection models and well-ordering of the model of *NFU*. It is clear that the resulting structure will be a model of TTU, with the expected special feature that the types are identical (and so the membership relations between successive types are identical as well).

It is the converse, that the consistency of TTU implies the consistency of *NFU*, which is less obvious. Suppose we have a model $\langle T, E, \mathtt{P1}, \mathtt{P2}, W \rangle$ of TTU.

The trick is to show that any increasing subsequence of types of this model can also be used to determine a model of TTU. Let f be a strictly increasing function from $\mathcal{N}$ to $\mathcal{N}$. We define a model $\langle T^f, E^f, \mathtt{P1}^f, \mathtt{P2}^f, W^f \rangle$ of TTU. T_i^f (type i of the submodel) will be simply $T_{f(i)}$. The definition of E_i^f is a little more complex: we need to develop a notion of "membership" of type $f(i)$ elements of our model in type $f(i+1)$ elements of our model: we know that $f(i+1) > f(i)$, but we do not necessarily have $f(i+1) = f(i) + 1$. The solution is to use the coding of elements of type $f(i)$ into $f(i+1) - 1$ as $(f(i+1) - f(i) - 1)$-fold iterated singletons. For each $x \in T_{f(i)}$, we write $\iota^{f(i+1)-f(i)-1}(x)$ to represent the iterated singleton of x (in the sense of the model) in $T_{f(i+1)-1}$ (note that this is x itself if $f(i+1) = f(i) + 1$). We define E_i^f as the set of all ordered pairs $\langle x, y \rangle$ such that $\iota^{f(i+1)-f(i)-1}(x) \, E_{f(i)-1} \, y \wedge (\forall z. z \, E_{f(i)-1} \, y \rightarrow (\exists w. z = \iota^{f(i+1)-f(i)-1}(w)))$. The new model regards all sets of $f(i+1) - f(i) - 1$-fold singletons (in the sense of the original model) in $T_{f(i+1)}$ as sets (it is easy to see that these correspond exactly to the sets of type $f(i)$ objects in the original model) and treats all other elements of $T_{f(i+1)}$ as urelements. Notice that the definition of E_i^f depends only on the values of $f(i)$ and $f(i+1)$; any model $\langle T^g, E^g, \mathtt{P1}^g, \mathtt{P2}^g, W^g \rangle$ of this kind which contains these two types as successive types will have the same membership relation between them. We define the sequences $\mathtt{P1}^f$, $\mathtt{P1}^f$, and W^f in the obvious way: e.g. $W_i^f = W_{f(i)}$. It is straightforward but tedious to show that this is in fact a model of TTU.

Now take any finite set of sentences $\mathcal{S}$ in the language of TTU. There will be a highest type n which occurs in any of these sentences. We can use $\mathcal{S}$ to define a finite partition of $[\mathcal{N}]^{n+1}$ as follows: the "compartments" of the partition into which an element A of $[\mathcal{N}]^{n+1}$ falls will be determined by the truth values of the sentences of $\mathcal{S}$ in models $\langle T^f, E^f, \mathtt{P1}^f, \mathtt{P2}^f, W^f \rangle$ where the image of $\{0, \ldots, n\}$ under f is A. Clearly there are no more than $2^{|\mathcal{S}|}$ compartments. By Ramsey's theorem, there is an infinite homogeneous set H for this partition. Let h be the strictly increasing map from $\mathcal{N}$ onto H. The model $\langle T^h, E^h, \mathtt{P1}^h, \mathtt{P2}^h, W^h \rangle$ will satisfy the scheme of typical ambiguity for sentences in $\mathcal{S}$. By compactness, it follows that the full scheme of typical ambiguity is consistent (because any finite subset of the scheme is consistent). By Specker's ambiguity theorem of the previous section, *NFU* is consistent.

It is a crucial feature of this construction that there is a natural way to define a "membership relation" between any two types i and j with $i < j$ in a model of *TTU*. There is no obvious way to "skip types" this way in a model of *TT* (and it is demonstrably impossible to do this in a model of *TT* which satisfies Choice). See our [12] for a discussion of this.

11 The Bootstrap to Untyped Foundations

At the moment, our metatheory is *TTU* and we have been reasoning in it about the object theories *TTU* and *NFU*. We now know (by dint of some hard reasoning, admittedly) that the consistency of *TTU* implies the consistency of *NFU*. Moreover, *NFU* is an extension of *TTU*: it extends *TTU* with the information that the types are all in fact one and the same domain.

On reflection (meta-meta-theoretic in nature!), we realize that we can extend our metatheory from *TTU* to *NFU* harmlessly. So (in at least a mathematical sense) we have arrived at foundations in *NFU* as desired.

12 The Meaning of Untyped Foundations

At this point we leave the task of mathematical development of our foundations and commence explicit philosophical reflection on the foundations (though we will not eschew further mathematical development). We do think that a basically self-contained exposition of the mathematics is of value for supporting the necessary reflection, which is why we have provided it.

Nothing we have done should necessarily have dispelled the impression that the development of *NFU* is a syntactical trick. What proponents of *ZFC* have that we do not have (so far) is a nice intuitive picture of what is going on in their favored approach to foundations: one starts with the empty set and runs through a series of stages indexed by the ordinals, at each step constructing all collections of objects constructed before that stage. This picture motivates not only *ZFC*, but a hierarchy of extensions of *ZFC* which is in principle impossible to formalize completely! An inessential modification of the intuitive motivation for *ZFC* gives an intuitive motivation for *ZFA* as well (one can start with a set of atoms, or allow the addition of sets of atoms at each stage).

Moreover, the motivation behind *TTU* might seem to be a variation of the same thing. One takes an arbitrary collection of objects to be type 0, then one takes all collections of these objects (plus some junk) to be type 1, all collections of type 1 objects (plus some junk) to be type 2, and so forth. As we have noted before, it is impossible to reconcile the idea of polymorphic collapse of a model of TTU with the idea that each type contains *every* subcollection of the previous type. To see this, consider the collection of all elements of a given type which do not belong to their polymorphic analogues in the next type; this would implement the Russell class! (this shows that the relation "x is the polymorphic analogue of y in type $\mathtt{type}(y) + 1$" cannot be representable inside the model of TTU being collapsed polymorphically).

As we have suggested at a couple of points in the mathematical development, this is not our understanding of TTU (and so we do not need to revise it when we make the transition to NFU). We suggested at the outset that we regard TTU as being fundamentally a theory of properties rather than a theory of classes (an intensional rather than an extensional theory). It is reasonable to suppose that some collections of an "arbitrary" nature may not be the extensions of any property of objects we recognize, and so may not be realized as sets. The extensional criterion of identity between sets may not be thought appropriate for a theory of properties, but it is certainly convenient for mathematical purposes, and we have shown how to arrange for it to hold by identifying properties with the same extension (in our interpretation of TTU in TTU_0 above).

The phenomenon of typical ambiguity in type theory that Russell and Quine noticed and from which Quine attempted to extrapolate is an intensional rather than an extensional feature of the theory; sets are being viewed as extensions of definable properties rather than "arbitrary" collections when the phenomenon of typical ambiguity is noticed.

NFU is an extensional theory in the sense that the criterion for identity between (instantiated) properties is extensional. It is not unreasonable even from an intensional and purely philosophical standpoint to maintain that properties which hold of exactly the same objects are the same (though this is not a commonly held view). Because NFU is extensional in this sense, it does make sense to refer to it as a theory of sets. But it is an intensional theory in the sense that the sets of NFU are considered as being extensions of properties rather than arbitrary collections (and we admit the possibility that some arbitrary collections may be not be extensions of sets because they are not the extension of properties we can specify).

Convention: We will refer to completely arbitrary collections here as "classes" rather than "sets". We will refer to the properties and relations whose extensions are sets as "natural" properties and relations.

It might be objected that our inclusion of choice in TTU is incompatible with regarding it as an "intensional" theory; it is often supposed that the purpose of the Axiom of Choice is precisely to allow the construction by "arbitrary" choices of sets which we cannot specify by a common property of their elements, making it incompatible with an "intensional" viewpoint. This is not necessarily the case. The Axiom of Constructibility ($V = L$) of Gödel is a statement motivated entirely by intensional considerations. The axiom $V = L$ is all about definability: every set in $ZF + V = L$ is definable from finitely many ordinal parameters, so is in a natural sense the extension of a property. But the "intensional" theory $ZF + V = L$ proves choice! Moreover, the objections to $ZF + V = L$ which are usually expressed are "extensional" in character: most set theorists doubt that the theory with $V = L$ captures the truth about "arbitrary" collections even of natural numbers!

The problem is to make intuitive sense of the stratification criterion of comprehension in NFU in a way that lets us see NFU as reflecting a view of the world with its own internal logic rather than a weird modification of the view of the world implicit in TTU. What is special about the properties with stratified definitions (i.e., why should we regard these properties as natural)?

The obvious shift from TTU to NFU is that the world of individuals with which we start (type 0) turns out to be the whole world! The properties and urelements of type 1 turn out to be identified (bijectively) with the individuals. Moreover, the details of the identification of type 1 with type 0 automatically handle all types at once; there is no opportunity to tinker so as to accommodate type 2 and higher types.

Speaking informally, what we have done is cause each natural property of individuals to be represented by an individual (the set with the extension determined by that property). We do not use all individuals to represent properties in this way; the ones we do not use are the urelements or atoms. This is analogous to a process with which we are all familiar: it resembles the process of assigning semantics to general terms (which might be regarded as common nouns, adjectives, or verbs) in a language. So we speak of the construction of our model of theory metaphorically as the construction of a kind of "language".

If we have a preexisting notion of "natural property" which we are trying to represent in this way (we suppose that a natural notion of ordered pair is provided so that we can identify natural relations with certain natural properties in the usual way), one relation which we cannot expect to be natural is the relation which holds between x and y iff x has the natural property represented by y: the semantic relations of the language can be expected to be "arbitrary relations" from the standpoint of the original notion of natural property (and relation) we started with. If the notion of natural property (and relation) satisfies reasonable closure properties under logical operations, this is provable (we can construct the paradox of heterological adjectives).

The stratification criterion does not disappoint us too much here: the relation $\in$ is not a set relation. But in the light of this consideration, the stratification criterion for comprehension and thus naturalness of properties looks uncomfortably strong: although $\in$ is not a set relation, many properties and relations defined in terms of $\in$ are required to be natural properties and relations by the stratification criterion.

Contemplation of the proof of the Meta-Theorem of finite axiomatization of NFU can be useful here. Lemma 1 expresses a closure property which we expect a reasonable notion of naturalness of properties to have: essentially, it ensures that properties and relations first-order definable in terms of natural properties and relations will be natural properties and relations. If we look at the proof of the Meta-Theorem, we see that the crucial fact is that inclusion (subsethood) is a natural relation (used to define the set relation E used to handle membership in the proof) and that natural relations on individuals induce natural relations on their singletons and vice versa (this is seen in the use of hierarchies of relations R_n induced by repeated application of singleton image in the definition of the modified formula ϕ', and in the use of set union at the last step to collapse a set of iterated singletons of elements of the desired set to the desired set).

We suggest an intuitive interpretation of the subset relation which is not semantic in character at all: the relation of subset to set can very reasonably be understood to be the restriction to sets of the relation of part to whole. This is not a semantic relation, and it is quite reasonable to suppose that it is a natural relation. Because $x \in y \equiv \{x\} \subseteq y$, it is clear that all that is needed now is an account of the singleton construction.

Further, our interpretation of what is going on in the construction of sets also suggests a metaphor for the singleton relation: the singleton set $\{x\}$ corresponds to the natural property of being x (the predicate of x-hood). We actually adopt a different metaphorical interpretation of $\{x\}$ which might seem surprising: we will call $\{x\}$ the "name" of x in the "language" we are constructing. In doing this, we are reversing a common maneuver for eliminating proper names a from formal logic: replacing proper names a with references to the predicate "being a" which applies only to a. Another way to see that it is reasonable to interpret $\{x\}$ as a proper noun standing for x is to interpret sets as common nouns (as if we were to associate the word "man" with the set of men); the proper noun "Peter" might then be seen to be associated with the set consisting only of Peter.

So we propose to reduce set theory to semantics and mereology (the study of the relation of part to whole). Something very like our reduction of set theoretical primitives to the subset relation (understood as the relation of part to whole) and the singleton relation is found in the works of David Lewis (notably [18]). We came up with this idea independently of Lewis (though we benefited subsequently from study of his development) and we provide an explanation of the role of the singleton relation where he treats it as something mysterious. Lewis analyzes standard foundations in *ZFC* (where such an analysis is equally valuable) rather than Quine-style set theory.

It is not to be expected that the relation between objects and their names will be a natural relation (and indeed we will prove below that it is not); it shares in the "arbitrariness" of the whole procedure of assigning referents to symbols in a language.

Our identification of the relation of subset to set with the relation of part to whole on sets allows us to give an intuitive picture of the properties of this "language". Singletons are disjoint (set-theoretically), so they are non-overlapping (mereologically). Singletons have no proper (nonempty) subsets which are sets, so they are "atomic" (at least among sets: they have no proper (and non-null) parts which are sets). We regard the empty set as the null part, about which we have no qualms (though we can fix this picture to satisfy such qualms by regarding the empty set as representing a particular atomic object and each singleton as the fusion of another atomic object with the atomic object representing the empty set).

We introduce a basic notion of mereology:

Definition: The *fusion* of all objects with property P is defined as the uniquely determined object which contains all objects with property P as parts and is part of every object which contains all objects with property P.

We make the mereological assumption that the fusion of all individuals with any natural property P is an individual. For the record, we also assume that the relation of part to whole is reflexive, antisymmetric, and transitive (a weak partial order).

A set is then presented to us as being the fusion of the names of all its elements. If P is a natural property, the property "being the name of something with P" is also a natural property (see the next paragraph), and by our mereological assumption there is an individual which is the fusion of all individuals with the latter property. This individual is taken to represent the property P. The role of the pairwise non-overlapping names (singletons) here is crucial. If we tried to use the fusion of the objects with property P to represent P, we would not be able to recover P unambiguously from the object representing it. For example, the fusion of all human beings and the fusion of all human cells are the same object (mod quibbles about tissue cultures and such; the point should be clear, though).

The crucial further property which makes everything work (as we can see by examining the proof of the Meta-Theorem) is that every natural property of individuals corresponds to a natural property of names and vice versa (and the same for relations – the axioms of singleton image and set union do this work in the axiom set) – even though the semantic relation between objects and names is not a natural relation. We are not saying that an individual and its name have the same properties – we are saying that for each natural property P, "being the name of something that has P" and "being the referent of something that has P" are also natural properties, and similarly for relations. The "world" of names is a "model" of the whole world for natural properties and relations: the structure consisting of the names is a kind of microcosm.

We suggest to the reader that this picture of the construction of a kind of "language" to express natural properties can be taken to be intuitively appealing; the obvious question about it is whether it is logically possible to do it. The development in the paper shows that if it is possible to model the many-sorted TTU (for which we have good intuition) it is possible to model the more problematic NFU. It should not be surprising that it takes nontrivial mathematical reasoning to show that this is possible.

13 The Intuitive Picture Summarized and the Types Revisited

We summarize our intuitive picture. We start with a world of individuals and notion of natural property and relation. Natural properties and relations are closed under first-order definability. Among the natural relations are the relations of equality, projections, the relation of part to whole, and a well-ordering of the universe. We attempt to develop a language in which each natural property and relation will be represented by an individual. We choose a relation of name to referent with the property that names are non-null and pairwise non-overlapping and that the natural properties and relations on names correspond precisely to the natural properties and relations on the corresponding referents. We assume that the fusion of all individuals with any given natural property is an individual. We represent each natural property by the fusion of the names of the individuals of which it holds; this will be an individual by an application of the last two assumptions. Recapitulation of the proof of the Meta-Theorem under a suitable interpretation establishes that we have a model of *NFU*.

Note that this picture is no more complicated than the intuitive picture of the cumulative hierarchy of types (recall that the intuitive picture of the cumulative hierarchy involves an account of the notion of the power set and of general ordinals). The disadvantage it has is that that it is less obvious that the picture can be realized (though it is not obvious to everyone that the cumulative hierarchy can be realized).

Notice also that this picture does not refer in any obvious way to types. Realization of this picture seems to be a reasonable thing to try in an untyped universe (or at least, one with only two types – a type of individuals and a type of natural properties). Yet we obtain an interpretation of full type theory. What is the meaning of the types if they are not to be taken (as they clearly cannot be taken here) to be disjoint sorts of object?

Once again, this is best discovered by looking at the proof of the Meta-Theorem. Specifically, we consider the relation of the formula ϕ' in that proof to the formula ϕ. In order to obtain the formula ϕ' which is entirely expressed in terms of natural relations from a stratified formula ϕ mentioning the non-natural relation $\in$, we need to replace talk of each object x mentioned in the proof with talk of its $(N - \mathtt{type}(x))$-fold singleton. In terms of our intuitive picture, we replace talk of x with talk of its $(N - \mathtt{type}(x))$-fold *name*.

When we look at an object in our intuitive picture, we can use it in various ways. We can use it directly. We can use it to refer to that (if anything) of which it is the name. And, further, we can iterate reference: we can look at the referent of its referent, the referent of the referent of its referent, and so forth. What the analysis in the previous paragraph reveals is that this is the function of the relative types in stratified comprehension from our untyped viewpoint. A stratified formula is one in which each variable x is used to refer to objects at the same iterated level of reference $((\mathtt{type}(x))$-fold reference) wherever it appears – knowledge of the relations between the different levels of reference of the same object requires knowledge of the reference relation itself, which is not a natural relation (we cannot expect to represent it internally to our language). Stratification is seen to be a natural property to expect of definitions of properties using the reference relation – those definitions which do not try to "diagonalize" on the reference relation succeed in defining natural properties. The reason that we are able to define any sets at all in terms of the reference relation has to do with the parallelism between natural properties of objects and natural properties of their names. We are also helped in defining sets in terms of membership ($\in$) by the part that the natural relation inclusion ($\subseteq$) plays in our understanding of membership. Moreover, types of individual variables are seen to be genuinely "relative" rather than absolute (if we increase the level of reference uniformly nothing changes, because of the perfect correspondence between properties of names and properties of referents).

The fact that there is a hierarchy of types is not surprising if we look at the fact that the world of "names" is postulated as a "microcosm" within the world of individuals: within the world of "names" as a model of the whole world, we would expect then to find a model of the world of "names" which can be viewed as an even smaller model of the whole world, and within that a yet smaller model of the world of "names", and so forth. If we take the first n of these models and the natural "reference relations" among them, we get a structure corresponding exactly to the natural model of TTU_n we presented in an earlier section! The full infinite sequence of nested models cannot be a set (we'll discuss this further in the next section).

14 Fitness for Mathematical Applications

NFU is fit for mathematical applications. An extended study of this can be found in my [13]; one can also look at Rosser's [24]; the development there is in *NF*, but it is readily adapted to *NFU* (with a primitive type-level ordered pair replacing the definable Quine pair used in *NF*).

Features of the implementation of mathematics in *NFU* which might be found appealing or instructive are the use of Frege's definition of the natural numbers (the number 3 is the set of all sets with three elements) and the Russell-Whitehead definitions of cardinal and ordinal number. It does not seem to be as widely known as it should be that these definitions do not in and of themselves lead to paradox. Of course, the fact that *NFU* has a universal set has the same instructive quality.

There are also features of the implementation of mathematics in *NFU* which may be found annoying or even pathological. Some of these can be discovered by looking at the resolutions of the paradoxes in *NFU*.

The Russell paradox is readily averted by the stratification criterion for comprehension: $x \notin x$ is not stratified, so we do not have to think about the embarrassing $\{x \mid x \notin x\}$.

The Cantor paradox results from applying Cantor's theorem $|A| < |\mathcal{P}(A)|$ (the cardinality of a set is strictly less than that of its power set) to the universal set. It is clear that $|\mathcal{P}(V)| \leq |V|$! The resolution lies in the fact that the form of Cantor's theorem provable in TTU is $|\mathcal{P}_1(A)| < |\mathcal{P}(A)|$: the set of one-element subsets of A is strictly smaller in size than the set of all subsets of A. We believe this, but we also tend to think that $|A| = |\mathcal{P}_1(A)|$. However, Cantor's theorem tells us that $|\mathcal{P}_1(V)| < |\mathcal{P}(V)| \leq |V|$; so it cannot be the case that $|V| = |\mathcal{P}_1(V)|$. We see that the singleton map is not a set function (from this it follows that the relation of name to referent in our intuitive picture is not a natural relation; we suggested that this would be the case).

We will later find occasion to use this

Definition: Cantor's theorem in its original form would hold if $|A| = |\mathcal{P}_1(A)|$ were true; we call sets A with this property *cantorian* sets. Of more interest to us are the *strongly cantorian* sets, which we define as those sets A such that the restriction of the singleton map to A is a set function.

The Burali-Forti paradox results from considering the order type of the natural order on the ordinals. The natural order on the (Russell-Whitehead) ordinal numbers is a set relation in *NFU* and in fact a well-ordering; thus it has an order type Ω. The Burali-Forti paradox results from applying the theorem of naive set theory (and *ZFC*) which asserts that the order type of the natural order on the ordinals less than an ordinal α is α. This would force Ω to sit past the end of the sequence of all ordinals, which is absurd! But this is not a theorem of TTU or *NFU*: the order type of the natural order on the ordinals less than α in TTU is an ordinal two types higher than α: if W is a well-ordering of type α, then $(W^\iota)^\iota$ (the order on double singletons induced by W) is the order type of the ordinals less than α. (We use the notation $T(\alpha)$ for the order type of W^ι, thus $T^2(\alpha)$ for the order type of $(W^\iota)^\iota$). Since the order type of the natural order on the ordinals less than Ω is obviously strictly less than Ω, which is the order type of *all* the ordinals, we have $T^2(\Omega) < \Omega$. T (and so T^2) respects the natural order on the ordinals, so repeated application of T^2 (or of T) to Ω produces an externally countable decreasing sequence of ordinals; this clearly cannot be a set, so T^2 (and thus T) is not a function. In some external sense, the ordinals of *NFU* are not well-ordered; this has been presented as a serious criticism of *NF*. The "sequence" $T^i(\Omega)$ of ordinals is a relic in *NFU* of the "hall of mirrors" in TTU; similar phenomena occur in the cardinal numbers.

The peculiar phenomena found in the resolutions of the paradoxes are certainly surprising to a student of standard set theory. However, our intuitive motivation for this style of set theory should already have suggested that some arbitrary collections of sets would be found not to be sets. The fact that the singleton map is not a function is quite natural in terms of our intuitive motivation. This allows V and $\mathcal{P}_1(V)$ (which are externally isomorphic) to have different cardinalities "internally" to *NFU*: in terms of our intuitive motivation, $|V|$ is the number of individuals and $|\mathcal{P}_1(V)|$ is the number of *names*: it is entirely reasonable in terms of our intuitive picture that the realm of names turns out to be "smaller than" (though externally isomorphic to) the universe. The fact that the ordinals are not externally well-ordered is another example of the same phenomenon: the definition of a well-ordering provides that any nonempty *subset* of the domain of the well-ordering has a least element, not that any *subclass* of the domain has a least element. The existence of a countable proper class is quite arresting, but it is particularly arresting to us because our intuition is trained by a system of set theory motivated by "limitation of size"; we believe that small collections should be sets. This is not part of the motivation of *NFU*.

We report from our own experience that working in *NFU* does allow one to develop some intuition about the properties one can expect sets to have in this system; in fact, we believe that once one understands what is going on, one finds that *NFU* is not nearly as far from *ZFC* in its outlook on the mathematical universe as one might suppose. This is the subject of the next section.

15 Mutual Reflections

In this section, we will discuss the relationships between the views of the mathematical universe embodied in *TTU*, *NFU*, and *ZFC*.

There is a nice interpretation of *NFU* in terms of the cumulative hierarchy of *ZFC*, which our goal of showing the autonomy of foundations in *NFU* did not allow us to present as our official consistency proof of *NFU*. If one takes a nonstandard model of (enough of) *ZFC* with an external isomorphism j between a rank $V_{\alpha+1}$ of the cumulative hierarchy with infinite index and a lower rank $V_{j(\alpha)+1}$ (this is usually obtained by providing an external automorphism of the entire model moving an ordinal downward, but an application below requires us to state the result with more generality) then one readily obtains an interpretation of *NFU*. The domain of the model of *NFU* will be the set V_α in the nonstandard model of set theory (here we use the notation V_α to stand for the rank indexed by α in the cumulative hierarchy of sets in *ZFC*; this should not be confused with the notation V^n for the set of n-fold singletons used above). The membership relation $x \in_{\texttt{new}} y$ of the model will be defined as $j(x) \in y \wedge y \in V_{j(\alpha)+1}$. Projection relations are available because α is infinite, and we can suppose ourselves provided with a well-ordering because the ambient theory has choice. We do not present a proof here that this is a model of *NFU*; the flavor of the proof is very similar to that of the proof of the Meta-Theorem above (the idea is to exhibit a transformation which eliminates the external function j from translations of stratified instances of comprehension, so that they can be seen to define sets in the underlying nonstandard model of set theory – see [5], pp. 68-9 for details). Work with this interpretation of *NFU* in *ZFC* (and similar theories) allows one to see that the viewpoint of *NFU* is not really profoundly different from that of *ZFC*.

It is also easy to interpret set theory in the style of Zermelo in *NFU*. The idea (due to Roland Hinnion in [10] – a treatment is found in [13]) is to interpret sets in Zermelo-style set theory as isomorphism classes of well-founded extensional relations with top elements: in other words, a set in the Zermelo-style set theory is represented by the isomorphism class of the membership relation restricted to its transitive closure (strictly speaking, the transitive closure of its singleton under the membership relation). The isomorphism classes of well-founded extensional relations make up a set, and there is a natural "membership" relation on these isomorphism classes (viewed as pictures of sets) which is also a set relation. If the whole domain of isomorphism classes is used to interpret set theory in this way, one always obtains a model of *ZFC* - Power Set in which Power Set is actually false (there is a largest cardinal). But if the ambient *NFU* is augmented with strong assumptions, restricting the domain of the interpreted set theory may yield a model of Zermelo set theory or *ZFC*. One sometimes restricts oneself to isomorphism classes of *strongly cantorian* well-founded extensional relations; in this case the domain of the interpreted Zermelo style set theory is a proper class (the predicate "strongly cantorian" is unstratified), but there are technical advantages to doing this. It may be instructive to see that Zermelo-style set theory can be understood as the theory of isomorphism classes of well-founded extensional relations, independently of any consideration of *NFU*.

The interpretation of *NFU* in *ZFC* that we have described can carried out naturally in Zermelo-style set theory as interpreted inside *NFU*! The point is that the model of Zermelo-style set theory one obtains inside *NFU* has an external endomorphism (isomorphism with a proper substructure of itself) which is definable in *NFU*, and can be used to define an external isomorphism between ranks of the cumulative hierarchy as above. For any well-founded extensional relation W with top, one can see that W^ι (the relation on singletons induced by W) is also a well-founded extensional relation with top, which is not necessarily isomorphic to W. If the isomorphism type of W is x, we denote the isomorphism type of W^ι by $T(x)$. T is easily seen to be an (external) endomorphism of the isomorphism types. T can be proved to map some rank of the cumulative hierarchy in the interpreted Zermelo-style set theory onto a lower rank. We then construct an interpretation (not a set model; there is no problem with Gödel's second incompleteness theorem here) of *NFU* in the set model of Zermelo-style set theory in the same way that *NFU* is interpreted in *ZFC* above. We think that it is definitely of interest that *NFU* reflects the strategy for construction of models of *NFU* in *ZFC* internally.

We look at the interpretation of TTU in NFU. We have observed above that TTU provides us with a natural model of TTU_n for each concrete natural number n. This interpretation gives us a natural model of TTU_n for each concrete natural number n in *NFU* (natural only in the sense that all subsets in *NFU* of any type are realized in the next higher type). Further, we can see by comparing the construction with terminology introduced during the proof of the Meta-Theorem that in these models type i is represented by $V^{(n-i)-1}$ (the set of $((n-i)-1)$-fold singletons) and membership of type i in type $i+1$ is represented by $E_{(n-i)-1}$ (where E is the relation introduced in Lemma 2). These models are all substructures of the same proper class model of a "type theory" with a top type but no bottom type. Type theory with types indexed by negative integers (or all integers) is consistent by a compactness argument (any proof in type theory uses only finitely many types, so any proof of a contradiction in the theory of negative types would immediately yield a contradiction in standard type theory – this is due to Hao Wang in [29]). This structure is not a set in *NFU* (if it were, we would fall afoul of Gödel's second incompleteness theorem, because the theory of this structure is as strong as TTU and so as *NFU* itself), and the sequences of its types and membership relations give new examples of countable proper classes. This is a better example than the descending sequence in the ordinals of how the "hall of mirrors" effect in TTU can be rediscovered in *NFU*. An elegant and surprising application of the properties of this structure in *NF*, which is also applicable to *NFU*, is found in [21].

Finally, we compare *NFU* and *ZFC* foundations in the light of our intuitive motivation for set theory using semantics and mereology, which is equally applicable to either theory.

The intuition behind the cumulative hierarchy looks perhaps a little less convincing (at least, one sees more clearly the enormity of the claim underlying this picture) when one presents it in terms of this interpretation of set theory. The thing which appears more remarkable in this interpretation is the construction of power sets at successor stages. Suppose one has constructed V_α. $V_{\alpha+1}$ is the set of all subsets of V_α; that is, it is the collection of all fusions of singletons of elements of V_α. Given the singletons, the construction of all the fusions representing sets of elements of V_α is nothing remarkable; one can quite reasonably adopt the view that any fusion of objects whatsoever is an object. But the construction of the singletons themselves should give us pause. V_α is a collection of objects which may overlap one another in very complicated ways: we implicitly claim that we can produce a collection of pairwise non-overlapping objects in one-to-one correspondence with any V_α. The intuition behind this must be that we have a truly inexhaustible supply of distinct atomic objects somewhere; this may not be an unreasonable assumption (there's no reason to think it leads to paradox) but it should give us pause.

The intuition behind TTU might be mistakenly taken to be the same as that behind *ZFC*: this is because one might suppose that we are simply building $V_{\omega+\omega}$ in *ZFC*. But the subtlety of the theory of types is that we make no assumptions about whether the objects at the next type are "new" or not. One does seem to be more or less forced to identify the finite well-founded sets in different types as being in some sense "the same" (or else one has to complicate the semantic relations in the intuitive picture), but what sorts of identifications may exist beyond that are open. One could suppose that each type $n+1$ properly extends type n (as in the interpretation in $V_{\omega+\omega}$) but one could equally well (and with some philosophical justification) suppose that each type $n+1$ is a proper subset of type n. Type 0 could be taken to be the universe at the outset – then one might suppose that the natural properties of the universe were coded in a small subset of the universe (in the sets of a model of *NFU*, say) and the natural properties of this subset of the universe were coded in a smaller subset still, etc.

The intuition behind *NFU* is seen to be quite daring (the nontrivial part is the idea that there is a domain of "names" natural properties on which reflect the natural properties on the universe precisely) but it has its own natural origin as a refinement of TTU, and a little mathematical work in TTU shows that it can be realized. We believe that we have adequately established that *NFU* can be understood as an autonomous view of the world of mathematics.

16 Extensions

A fundamental (and philosophically appealing) characteristic of the foundational picture behind *ZFC* is that it is "self-extending". If one understands the iterative hierarchy of ranks in *ZFC*, it seems entirely reasonable that there would be an inaccessible cardinal, and this intuition continues to support one (perhaps with increasing doubts) as stronger and stronger large cardinal axioms are considered.

Foundations in *NFU* have the same characteristic. Formal features of the theory suggest extensions of the theory which turn out to be consistent (on reasonable assumptions) and usually surprisingly strong.

Natural extensions of *NFU* tend to involve the notions of *cantorian* and *strongly cantorian* sets introduced above. We will review a few extended versions of *NFU* to get a flavor of what is going on.

The reason why "strongly cantorian" proves to be an important notion in formulating extensions of *NFU* has to do with the fact that stratification restrictions on the formation of sets $\{x \mid \phi\}$ can be avoided to some extent where variables restricted to strongly cantorian sets are involved. If A is a strongly cantorian set, then there is a function ι_A such that $\iota_A(x) = \{x\}$ for all $x \in A$, and also a function ι_A^{-1} such that $\iota_A^{-1}(\{x\}) = x$ for all $x \in A$. It follows that we can lower the relative type of any variable x restricted to A by replacing it with the equivalent term $\iota_A^{-1}(\{x\})$, and raise its relative type by replacing it with the equivalent term $\bigcup(\iota_A(x))$ (this notationally compact formulation for type-raising only works if all elements of A are sets, but there is no essential difficulty in referring to "the element of $\iota_A(x)$" in case x might be an urelement). Either of these operations can be iterated, so the type of a variable restricted to a strongly cantorian set can be reset to any desired value; the types of such variables can be ignored in determining stratification.

An extension of *NF* which recommended itself to Rosser in [24] and is equally appropriate for *NFU* is embodied in the following "obviously true":

Axiom (Rosser): For each natural number n, $\{1, \ldots, n\}$ has n elements.

Rosser called this the Axiom of Counting; it expresses our intuitive notion that counting can be effected by putting finite sets into one-to-one correspondence with initial segments of the natural numbers in a sensible way. It might seem that one could prove this as a theorem by induction on n, but it happens that the condition $\{1, \ldots, n\} \in n$ is not stratified. It is sufficient on the basis of the discussion above to assume that $\mathcal{N}$ is strongly cantorian – then the set of all natural numbers satisfying the property in the Axiom of Counting would exist and would have to be all of $\mathcal{N}$. It turns out that in fact the Axiom of Counting is equivalent to the assertion that $\mathcal{N}$ is strongly cantorian (and that there are models of *NFU* in which it does not hold).

The Axiom of Counting has surprisingly strong consequences in set theory. *NFU* proves the existence of $\aleph_n$ for each concrete n, but does not prove the existence of $\aleph_\omega$. *NFU* + Counting proves the existence of $\aleph_{\aleph_n}$ for each n, which is a surprising gain in strength.

An innocent-seeming axiom proposed in [11] by C. Ward Henson in the context of *NF* (in a slightly different form) is the

Axiom of Cantorian Sets: Each cantorian set is strongly cantorian.

This implies Counting because *NFU* proves that $\mathcal{N}$ is cantorian. The axiom says that any set which is the same size as its image under the singleton map actually supports the restriction of the singleton map as a set function, which seems reasonable.

Theorem (Solovay): *NFU* + Axiom of Cantorian Sets has the exact consistency strength of *ZFC* + the scheme which asserts for each concrete natural number n that there is an n-Mahlo cardinal.

This is a surprising level of strength! We prove in [14] that *NFU* + Axiom of Cantorian Sets really does prove that there are n-Mahlo cardinals for each concrete n (Solovay showed that they existed in an appropriate version of the constructible universe).

Our intuitive picture tells us that a strongly cantorian set is a set on which we are given the relation of name to referent as a natural relation. We have seen that this allows the relative type of a variable restricted to this set to be ignored. It may suggest the stronger idea that "arbitrary" subclasses of strongly cantorian sets might be sets. We present an axiom along these lines.

Definition: A strongly cantorian ordinal is the order type of a strongly cantorian well-ordering (a strongly cantorian ordinal will not itself be a strongly cantorian set – it will be a very large set because it is an isomorphism class).

Axiom of Small Ordinals: For any formula ϕ there is a set A such that the elements of A which are strongly cantorian ordinals are precisely the strongly cantorian ordinals x such that ϕ; in other words, the intersection of any definable class with the strongly cantorian ordinals is the intersection of some set with the strongly cantorian ordinals.

Theorem (Holmes, closely following Solovay): The consistency strength of *NFU* + Counting + Small Ordinals is exactly that of *ZFC* - Power Set + "there is a weakly compact cardinal".

An intuitively reasonable extension of *NFU* once again gives considerable consistency strength (note that this theory is stronger than *ZFC*, in spite of the omission of Power Set, because a weakly compact cardinal is inaccessible). For the proof, see Solovay's [25] and my refinement in [14]. A refinement of this theory (described in [13]) gives the strength of second-order *ZFC* with a measure on the proper class ordinal (as proved in our pending [14]).

The Axiom of Small Ordinals suggests that the sequence of strongly cantorian ordinals is to be identified with the ordinals of *ZFC*. Comparison of the definition of the ordinals in *ZFC* with that in *NFU* is interesting. The class of von Neumann ordinals cannot be a set in any set theory (it is a paradoxical totality like the Russell class). One cannot even prove the existence of the von Neumann ω in *NFU* (its definition is unstratified). However, the existence of infinite von Neumann ordinals is consistent with *NFU*. Any von Neumann ordinal must be strongly cantorian, and it is easy (using permutation methods – see [5]) to convert any model of *NFU* to a model of *NFU* with a von Neumann ordinal corresponding to each strongly cantorian Russell-Whitehead ordinal. This suggests that it is the "limit" of the proper class of strongly cantorian ordinals rather than a "big" set ordinal like Ω which corresponds (as it were) to Cantor's Absolute.

The purpose of this section is to suggest that the approach to foundations based on *NFU* may have considerable mathematical power. The axioms adjoined to *NFU* here are natural statements to consider in the context of *NFU* (and its intuitive picture); one could consider systems like *NFU* + "there is an inaccessible cardinal", but these would not witness the autonomy of *NFU* foundations, since "there is an inaccessible cardinal" is an axiom suggested by the *ZFC* viewpoint. Notice that *NFU* + Axiom of Cantorian Sets does prove that there is an inaccessible cardinal (and more).

17 Review of Philosophical Objections

In this section, we review common objections to *NFU* of a philosophical nature.

There are objections to *NFU* which apply to *ZFC* as much as they apply to *NFU*, and so are not directly relevant to our purpose here. Both theories assume the axiom of choice, for which we will not apologize. Both theories are classical rather than constructive. There has been a little work on intuitionistic *NF* (which is not known to be consistent – see for example [4]) and as far as we know none on intuitionistic *NFU*. Both theories are impredicative. In this connection it might be worth noting that if one restricts the comprehension scheme of *TTU* to prevent the occurrence of variables of higher type than that of the set being defined, and further assumes strong extensionality, it is known that the type structure of the resulting theory *TTI* can be collapsed to obtain an untyped theory *NFI*. A further refinement of the restriction on comprehension in *TT*, forbidding bound variables of the same type as the set being defined in instances of comprehension, gives a truly predicative type theory *TTP* which can be collapsed to an untyped theory *NFP*. The untyped theories are no stronger than the related predicative type theories (as long as Infinity is assumed in the typed theories: *NFI* is exactly as strong as second-order arithmetic, while *NFP* is weaker than first-order arithmetic but does prove Infinity), but it is not clear that they can really be viewed as predicative theories from a philosophical standpoint: they do admit self-membered sets, for example. These results are discussed in Marcel Crabbé's paper [2].

The theory *NFP* is finitely axiomatized by the axiom set we give for *NFU*, modified by the addition of strong extensionality and the omission of Set Union. The proofs of Lemmas 1 and 2 do not use the set union axiom. The proof of the Meta-Theorem requires for success that $N = \texttt{type}(x)$; parameters of type one higher than that of x (but not bound variables) are harmless; atomic sentences involving these parameters do not need to be translated because they are already in a form which can be handled by Lemma 1. The intuitive picture we gave for *NFU* works for *NFP* as well, with the modification that we assume that all properties of referents are reflected by properties of names, but we do not assume that all properties of names are reflected by properties of referents (strong extensionality corresponds to the further assumption that all individuals are fusions of names).

Some object to the urelements. Forster has gone so far in his otherwise excellent recent survey [6] of *NF* as to suggest that to adopt *NFU* as a foundation is to "betray set theory". We fail to see any particular philosophical appeal in restricting oneself entirely to pure sets. Zermelo's original formulation of set theory admitted atoms (and also admitted non-well-founded sets). The step from *ZFA* to *ZF* is technically easy – one can define the class of pure sets (the definition involves recursion on the membership relation) and restrict one's attention to pure sets thereafter (and this has technical advantages). In *NFU* no such restriction to pure sets is possible. The definition of the class of pure sets is inherently unstratified – no construction involving recursion on the membership relation can be expected to succeed in *NFU*.

It is surprising that the existence of urelements is actually *provable* in *NFU* (using Choice). But it is no more than surprising; it doesn't cause any fundamental problems for the user of the theory. The natural methods of constructing models of *NFU* in *ZFC*, *TTU* or *NFU* itself all involve the creation of urelements anyway. Of course Specker's disproof of Choice in *NF* (logically equivalent to the disproof of strong extensionality in *NFU* with Choice) is a serious problem for one who supports foundations in *NF* – which we do not. (See [26] for this result, or [5], p. 50; it may also be instructive to look at my proof that there are atoms in [13], which makes use of the Axiom of Counting to simplify Specker's argument).

Another objection to the urelements is that they are a large collection of apparently indistinguishable objects with no obvious function in the theory (except avoidance of contradiction). This is by contrast with *ZFC*, the standard model of which is "rigid" (all objects in a model of *ZFC* are in some sense distinguishable from one another). We observe that a taste for rigid structures is relatively recent. If we look at the early history of mathematics, we see arithmeticians studying a rigid structure, but geometers studying a homogeneous structure: the points of Euclidean space are indistinguishable from one another. The implementation of Euclidean space as R^2 in *ZFC* is "rigid", but it is not clear that Euclidean space really comes with a preferred set of Cartesian axes built in. The distinct elements of type 0 in Russell's theory of types have a similar homogeneous character.

Philosophers should notice that there do seem to be non-sets in the real world, and even that the world seems to support some homogeneity (as between points of space and moments of time).

There is a large class of objections to phenomena in *NFU* which reflect the fact that it is not motivated by the "limitation of size" doctrine. Because *NFU* is a set theory with a univeral set, it is possible for a set (e.g., V) to have a proper subclass (e.g., the Russell class). The fact that *NFU* further admits countable proper classes is quite arresting. But the reasons that these proper classes are not sets are clear in terms of the intuitive picture of the theory we have presented (they involve recursion along the "arbitrary" relation of name to referent). The failure of the ordinals of *NFU* to be externally well-ordered is another phenomenon of the same kind (something like the Axiom of Small Ordinals can recover sensible behavior on the part of small ordinals, while our intuitive motivation suggests that the behavior of "big" ordinals can be expected to be odd).

All of these objections are answered at least to some extent by the observation that the underlying motivation of *NFU* does not require (indeed forbids) us to assume that each arbitrary extension is realized by a set. There are other set theories, notably the set theory *IST* used as a foundation for nonstandard analysis, which admit (and even exploit) the possibility of proper subclasses of sets.

Mathematicians seem particularly uncomfortable with the idea that A and $\mathcal{P}_1(A)$ can be of different sizes in *NFU*. In our experience, this particular phenomenon is something one can get used to; one can develop an intuitive feel for the difference between set bijections and those that involve a shift of type.

Those who think that the identification of successive types in models of type theory is an unnatural maneuver may be given pause by the fact that *all* infinite models of TTU_3 (including models of TT_3) satisfy the ambiguity scheme (of course, we can only define $\phi \equiv \phi^+$ for sentences ϕ which do not mention the top type). This implies that the theories NF_3 and NFU_3 in which only those instances of the comprehension scheme are assumed which can be stratified with three types have lots of models. NFU_4 is equivalent to *NFU* (and also $NF_4 = NF$). (These are results of Grishin (see [8]), but his papers are not very accessible: it is better to look at [5], p. 65). There are nice consistency proofs for both *NFI* and *NFU* (due to Marcel Crabbé) which exploit the ambiguity of models with three types.

Finally, there is the objection that attention to stratification is somehow an unnatural constraint on mathematical work. We think that in fact most mathematical constructions are naturally stratified, mod a universal tendency to confuse objects with their singleton sets (which is of course deplorable from the *NFU* standpoint). An occasional but not uncommon inconvenience in mathematical constructions in *NFU* is that one finds oneself explicitly using the singleton construction or an operation like $\mathcal{P}_1$ or the T operations on cardinal or ordinal numbers to fix a stratification problem. On the other hand, the most commonly used unstratified construction, the construction of the von Neumann ordinals (including natural numbers) is quite unnatural by comparison with the Russell-Whitehead constructions used in *NFU* (one feels this especially strongly when teaching it to students), though it is undeniably convenient.

We feel that this objection may have some merit: Zermelo style foundations may be slightly simpler on a technical level than Quine-style foundations. But Quine-style foundations are technically feasible (not just usable in principle). Our experience indicates that this approach to foundations can be used in practice.

In connection with the "naturalness" of the stratification criterion, it is natural to mention Forster's elegant theorem to the effect that the stratified predicates are exactly those sentences which are invariant under the redefinition of the membership relation using "setlike" permutations in any extensional structure for the language of set theory, which strongly suggests that stratification is a natural property for reasons not obviously related to those given here (we don't consider it to be entirely unrelated). The details can be found in [5], p. 94.

Our aim in this paper is not to suggest any change in mathematical practice. *ZFC* foundations are quite satisfactory, in our opinion. But if one wants to understand what foundations are and what they do for us, it is useful to be aware that there are other possible approaches.

References

[1] M. Boffa, "The consistency problem for *NF*", *Journal of Symbolic Logic* 42 (1977), pp. 215-220.

[2] Marcel Crabbé, "On the consistency of an impredicative subsystem of Quine's *NF*," *Journal of Symbolic Logic*, 47 (1982), pp. 131–36.

[3] Marcel Crabbé, "On *NFU*", *Notre Dame Journal of Formal Logic* 33 (1992), pp 112-119.

[4] Daniel Dziergowski, "Le théorème d'ambiguïté et son extension à la logique intuitionniste" (in French), doctoral thesis, Catholic University of Louvain, 1993.

[5] T. E. Forster, *Set theory with a universal set, an exploration of an untyped universe*, 2nd. ed., Oxford logic guides, no. 31, OUP, 1995.

[6] T. E. Forster, "Quine's NF, 60 years on", *American Mathematical Monthly*, vol. 104, no. 9 (November, 1997), pp. 838-845.

[7] Foundations of Mathematics mailing list: Steve Simpson, moderator, home page with archive at `http://www.math.psu.edu/simpson/fom/`

[8] V. N. Grishin, "Consistency of a fragment of Quine's *NF* system," *Sov. Math. Dokl.*, 10 (1969), pp. 1387–90.

[9] T. Hailperin, "A set of axioms for logic". *Journal of Symbolic Logic*, 9 (1944), pp. 1–19.

[10] Roland Hinnion, "Sur la théorie des ensembles de Quine", Ph.D. thesis, ULB Brussels, 1975.

[11] C. W. Henson, "Type-raising operations in *NF*", *Journal of Symbolic Logic* 38 (1973), pp. 59-68.

[12] M. Randall Holmes, "The equivalence of *NF*-style set theories with "tangled" type theories; the construction of ω-models of predicative *NF* (and more)", *Journal of Symbolic Logic* 60 (1995), pp. 178-189.

[13] M. Randall Holmes, *Elementary Set Theory with a Universal Set*, volume 10 of the Cahiers du Centre de logique, Academia, Louvain-la-Neuve (Belgium), 1998.

[14] M. Randall Holmes, "Strong axioms of infinity in *NFU*", to appear in the Journal of Symbolic Logic.

[15] M. Randall Holmes and J. Alves-Foss, "The Watson Theorem Prover", preprint, available at the Watson home page `http://math.boisestate.edu/~holmes/proverpage.html`

[16] M. Randall Holmes, "A strong and mechanizable grand logic", to appear in J. Harrison and M. Aagaard, *Theorem Proving in Higher Order Logics: 13th International Conference, TPHOLs 2000*, Lecture Notes in Computer Science, vol. 1869, Springer-Verlag, 2000, pp. 283-300.

[17] Ronald Bjorn Jensen, "On the consistency of a slight (?) modification of Quine's 'New Foundations'," *Synthese*, 19 (1969), pp. 250–63.

[18] David Lewis, *Parts of Classes*, Basil Blackwell Ltd., Oxford, 1991.

[19] Saunders Mac Lane, *Mathematics, Form and Function*, Springer-Verlag, 1986.

[20] A. R. D. Mathias, "Notes on Mac Lane set theory", to appear in Annals of Pure and Applied Logic.

[21] S. Orey, "New Foundations and the axiom of counting," *Duke Mathematical Journal*, 31 (1964), pp. 655–60.

[22] W. V. O. Quine, "New Foundations for Mathematical Logic," *American Mathematical Monthly*, 44 (1937), pp. 70–80.

[23] W. V. O. Quine, "On ordered pairs". *Journal of Symbolic Logic* 10 (1945), pp. 95-96.

[24] J. Barkley Rosser, *Logic for Mathematicians*, second edition, Chelsea, New York, 1973.

[25] Robert Solovay, "The consistency strength of *NFUB*", preprint, available through logic e-prints on the WWW.

[26] E. P. Specker, "The axiom of choice in Quine's 'New Foundations for Mathematical Logic'," *Proceedings of the National Academy of Sciences of the U. S. A.*, 39 (1953), pp. 972–5.

[27] E. P. Specker, "Typical ambiguity," in *Logic, methodology, and philosophy of science*, ed. E. Nagel. Stanford, 1962.

[28] Alfred Tarski and Steven Givant, *A Formalization of Set Theory Without Variables*, American Mathematical Society, Providence, 1988.

[29] Hao Wang, "Negative types", *Mind* 61 (1952), pp. 366-368.