Modelling fragments of Quine's "New Foundations"

M. Randall Holmes

In 1935, W. V. O. Quine introduced the set theory usually referred to as "New Foundations" (after the title of [1], in which it was first described), abbreviated NF. NF is a first-order theory with equality and membership, and the following non-logical axiom and axiom schema:

(Ext) $(\forall z)(z \in x \Leftrightarrow z \in y) \Rightarrow x = y$

(Comp) $(\exists y)(\forall x)(x \in y \Leftrightarrow P)$,
for each formula P which is "stratified" and does not contain y free.

A formula P is *stratified* iff it is possible to assign an integer "type" to each variable appearing in P in such a way that the types of variables u and v in subformulæ $u = v$ of P are the same and the type of the variable v is the successor of the type of the variable u in each subformula $u \in v$ of P. The axiom schema *(Comp)* can be replaced by a finite list of axioms (see [2], for instance). The set $\{x : P\}$ can be shown to exist using *(Comp)* precisely if the variables in P can be assigned type indices in such a way that the existence of $\{x : P\}$ becomes an instance of the axiom scheme of comprehension of the simple theory of types, but NF is an untyped theory. Note that definite descriptions $(\tau x)(P)$, meaning *the unique x such that P, or the empty set if there is no such x*, can be added to the language of NF; a term $(\tau x)(P)$ is assigned the same type as x for purposes of stratification.

An obvious difference between NF and the usual set theories is that, since $x = x$ is a stratified formula, the set $V = \{x : x = x\}$ exists. Clearly, $V \in V$ is a true statement, so one cannot adjoin an axiom of foundation to NF. NF has other peculiarities: notably, it is possible to disprove the Axiom of Choice (and so prove the Axiom of Infinity) (see [3]). The consistency of NF remains an open question (see [4]), but several interesting fragments of NF are known to be consistent. We discuss a technique for modelling fragments of NF which can be used to model the fragment NFU (NF with ur-elements) of Jensen, introduced in [5], and the fragment NFI ("mildly impredicative" NF) of Crabbé, introduced in [6]. Our method uses models with automorphisms; the use of models of set theory with automorphisms to model NFU was introduced explicitly by Maurice Boffa in [7] (the basic idea of Jensen's original construction is similar).

NFU is presented in a form suggested by Quine in his remarks accompanying [5]. The language of NFU is the language of NF with the addition of a sethood predicate. The axiom *(Ext)* is replaced by an axiom *(Ext′)* and an axiom *(Nonset)* is adjoined, as follows:

(Ext′) $\mathrm{Set}(x) \wedge \mathrm{Set}(y) \wedge (\forall z)(z \in x \Leftrightarrow z \in y) \;\Rightarrow\; x = y$

(Nonset) $\neg\mathrm{Set}(x) \Rightarrow \neg y \in x$

(Comp) As in NF.

NFU has the same powerful axiom of comprehension as NF; it differs from NF in having objects which are not sets, which we call ur-elements, which share the same extension with one another and with the empty set. NFU is consistent and weaker than arithmetic; it is consistent with the Axiom of Choice and can be extended with axioms of infinity so as to be powerful enough to found as much of mathematics as one might wish; our method of model construction will make this clear. Note that NFU, like NF, has a universal set, so cannot satisfy an axiom of foundation.

We work in ZFC. Let X be a transitive set. Define X_0 as X, $X_{\alpha+1}$ as the power set of X_α for each ordinal α, and X_λ as the union of the X_α's for $\alpha < \lambda$. Let $\emptyset$ be the empty set; let U be a nonstandard model of a $\emptyset_\beta$ which contains X_α for each $\alpha < \beta$, having an automorphism j which moves an ordinal α to a smaller ordinal $j(\alpha)$. For any $x \in U$, let x' abbreviate $j(x)$. We show how to interpret the object interpreted as X_α in U as a model of NFU.

Let $\in'$ stand for the membership relation in the nonstandard model U. For x and y in the set interpreted as X_α in the model U, define $x \in y$ as $x' \in' y \wedge y \in' X_{\alpha'+1}$. We claim that the X_α of the model U with the relation $\in$ is a model of NFU. Note first that *(Ext′)* and *(Nonset)* are easily seen to be satisfied: define $\mathrm{Set}(x)$ as $x \in' X_{\alpha'+1}$ —each object not in $X_{\alpha'+1}$ is interpreted as having no elements; if two objects are in $X_{\alpha'+1}$ and have the same "elements" under $\in$, it is easy to see that they have the same elements under the membership of U and so are equal.

We verify that *(Comp)* is satisfied. Let P be a stratified formula in the language of NFU. We know how to translate P into a formula P' in the language of U with references to the automorphism; we restrict each bound variable to X_α and replace each $x \in y$ with $x' \in' y \wedge y \in' X_{\alpha'+1}$. Choose an assignment of types to the variables of P which assigns 0 to the variable x; by applying the automorphism to the atomic subformulæ of P', we can arrange for each variable y of type n to appear everywhere with exactly $-n$ applications of the automorphism. This does not change the conditions for satisfaction of P'. We can then replace $j^n(y)$ with y wherever it occurs, restricted to $j^n(X_\alpha)$ instead of X_α. This does not change the conditions for the satisfaction of P', and removes all occurrences of the automorphism except in constants (iterated images of X_α and $X_{\alpha'+1}$ under j) and parameters. It follows that there is a set $A = \{x \in' X_\alpha \; : \; P'\}$ in U, and the set A' (not A) has as interpreted "elements" exactly those objects x in X_α of which P holds in the sense of the model of NFU (note that we would have obtained A' following the procedure above if we had assigned x the type -1 instead of 0). This procedure cannot be carried out for unstratified P, because it is not in general possible to eliminate the external automorphism from the translation P' in this case. This argument is an adaptation of the technique of Specker's [8] relating models of NF to models of the theory of types with type-shifting automorphisms.

We construct the natural numbers inside NFU. 0 is $\{\{\,\}\}$, the set whose sole element is the empty set (the unique **set** with no elements, as opposed to the ur-elements). If n is any object, we define $n+1$ as

$$n+1 \stackrel{\text{def}}{=} \{x \; : \; (\exists y)(\exists z)(y \in n \wedge z \notin y \wedge (\forall w)(w \in x \Leftrightarrow w \in y \vee w = z)\},$$

the set of all disjoint unions of elements of n and singletons. We define 1 as $0+1$, 2 as $1+1$, etc, thus defining each concrete natural number n as the set of all sets with n elements. We can define N, the set of all natural numbers, as

$$N \stackrel{\text{def}}{=} \{n \; : \; (\forall a)(0 \in a \wedge (\forall y)(y \in a \Rightarrow y+1 \in a) \;\Rightarrow\; n \in a)\},$$

the intersection of all inductive sets which contain 0. All $n > 0$ are very large collections, illegitimate in set theories of the usual type, but observe that each concrete n is interpreted in the model U by the set of all subsets of $X_{\alpha''}$ with n elements, a subset of $X_{\alpha'}$ —in terms of the model U, these are well-founded collections. If there are nonstandard natural numbers which are moved by the automorphism, peculiar things will happen. A subset A of X_α with n elements in the sense of the model U will be represented in NFU by the set A' with n' elements; the collection representing n will contain the objects B'' for each B with n elements. The set of all natural numbers less than n will be represented by $\{0, \ldots, n'-1\}$, a set with n' elements; the number of numbers less than n is n', and if n is moved by the automorphism, the number of numbers less than n will be distinct from n. This is not a contradiction: the formula *the set* $\{0, \ldots, n-1\}$ *belongs to* n is not stratified, so does not define a set; the set of natural numbers is defined as the intersection of all inductive sets, and this only enables us to carry out induction on properties which define sets —the inductive argument which shows that there are n numbers less than n is not valid in NFU. However, if no natural number is moved by the automorphism of the model U, our model of NFU satisfies the sentence $\{0, \ldots, n-1\}$ *belongs to* n *for each natural number* n, which is known as Rosser's Axiom of Counting, and strengthens NFU considerably. If the ordinal α used in the construction of our model is finite, the Axiom of Counting is obviously false; in this case the universe of the model of NFU is finite and its cardinality $|X_\alpha|$ is clearly moved to the smaller cardinality $|X_{\alpha'}|$ by the automorphism. In fact, the Axiom of Counting implies that the universe is infinite; the converse is not true —in a model U in which a natural number $n \neq n'$, we can still choose α infinite and obtain a model of NFU with the Axiom of Infinity (NFU + Infinity). We usually restrict our attention hereafter to models of NFU + Infinity. We give a definition of the ordered pair on X_α valid in the model U for infinite α:

$$(A, B) \stackrel{\text{def}}{=} \begin{array}{l} \text{the union of the set } \overline{A} \text{ obtained by replacing each natural number} \\ \text{appearing as an element of an element of } A \text{ with its successor} \\ \text{and the set } \overline{\overline{B}} \text{ obtained by adding 0 as an element of each element of } \overline{B}. \end{array}$$

This definition of the ordered pair is due to Quine, and has the property that it allows one to interpret all elements of X_α as pairs. We carry this pair over to our models of NFU. We prefer not to use the interpretation of the Kuratowski pair $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$ in NFU, because the object interpreting $\{\{x\}, \{x, y\}\}$ is $\{\{x''\}, \{x'', y''\}\}$, which is two types below x and y in any set definition; (x, y) has the same type as x and y. In NFU without Infinity, one has no choice but to use the Kuratowski pair; it must also be noted that the type-level pair (x, y) must be treated as a primitive notion in NFU + Infinity, since we cannot define it internally (although it is possible to interpret NFU + Infinity + existence of (x, y) in NFU + Infinity, by taking as objects of the former theory the "sets of sets" in the latter theory, on which the Quine pair can be defined, and redefining membership in such a way that "sets of sets of sets" have their original extensions, and all other sets of sets become ur-elements). With a type-level pair, we can develop the theory of relations and functions in NFU + Infinity along the usual lines (it is not much more difficult to do this with the Kuratowski pair, but it can be annoying that the type-differential between a relation and its domain is three instead of one).

We observe that there is an axiom of some interest which holds in our models of NFU but is not a theorem of NFU. Observe that the map which takes singletons of elements of $X_{\alpha''}$ to their elements exists in our model. In the sense of U, this function takes each $\{x''\}$ to x'', but in the sense of NFU, it takes each $\{x\}$ to x' —it takes singletons of objects to their images under the automorphism. We call this map Endo and abbreviate Endo($\{x\}$) as Endo$\{x\}$; we can then state the following:

Axiom of Endomorphism:

> *There is a map* Endo, *a bijection from the set of singletons into the set of sets, such that if A is a set,* $\text{Endo}\{A\} = \{\text{Endo}\{B\} \; : \; B \in A\}$.

This is an assertion easily verified by examining the definition of membership in the model of NFU. The Axiom of Endomorphism cannot hold in NF; we have shown this in [9] (so we can prove the existence of ur-elements in NFU + Endomorphism).

The Cantor theorem that the cardinality of the power set of A is greater than the cardinality of A for each set A is clearly false in NFU; consider $A = V$. In terms of the model U, observe that the "real" power set of B' is the object interpreted in NFU as the power set of B, while B' has the same cardinality as the object interpreted as the set of singletons of elements of B (the set of all objects $\{b'\}$ for $b \in B$ —Endo provides the bijection between B' and the interpreted set of singletons of elements of B). It is in fact a theorem of NFU that the cardinality of the power set of a set A is greater than the cardinality of the set of one-element subsets of A; the latter is not in general the same as the cardinality of A (consider 1, the set of all singletons, which has the cardinality of $X_{\alpha''}$ in U, while V, the universe, has the much larger cardinality of $X_{\alpha'}$ in U; the map taking each x to its singleton $\{x\}$ has an unstratified definition and does not exist in NFU). Sets A with the same cardinality as the set of one-element subsets of A are called *Cantorian* sets, and satisfy the Cantor theorem; even better are the *strongly Cantorian* sets, which have the property that the map taking x to $\{x\}$ for each x in A exists. N is provably Cantorian in NFU + Infinity; the assertion that N is strongly Cantorian is equivalent to the Axiom of Counting. Strongly Cantorian sets are of special interest because a variable restricted to a strongly Cantorian set A can have its "type" in a set definition $\{x \in A \; : \; P\}$ freely raised or lowered by exploiting the availability of the restricted singleton map; subsets of strongly Cantorian sets can be defined in unstratified ways. In our model construction, a set is Cantorian if it is fixed under the automorphism and strongly Cantorian if each of its elements is fixed under the automorphism (the converses are not necessarily true). Note that if all elements of the set X are fixed under the automorphism of U in the construction, X will be interpreted as a strongly Cantorian set; this is a technique for building models of NFU satisfying strong "axioms of infinity". For example, taking X countable with each element fixed under the automorphism gives a model of the Axiom of Counting, and taking X a model of ZFC with each element fixed under the automorphism gives a model of NFU containing a strongly Cantorian inaccessible cardinal.

In a model of NFU constructed by our technique, it is possible to use the function Endo to define the membership of the underlying X_α: $x \in' y$ in X_α iff $x' \in' y'$ in X_α iff $x \in y' = \text{Endo}\{y\}$ in NFU. We refer to the relation $x \in' y$ defined as $x \in \text{Endo}\{y\}$ as *type-level membership* in NFU + Endomorphism. The theory of "type-level membership" in NFU + Endomorphism has the following properties:

- if for all z, $z \in' x$ iff $z \in' y$, it follows that Endo$\{x\}$ and Endo$\{y\}$ are sets with the same elements, and so are equal;
- if we abbreviate Endo$\{x\}$ as $j(x)$, it is easy to see that $j(x) \in' j(y)$ iff $x \in' y$, $j(x) = j(y)$ iff $x = y$, and $x \in' V$ iff $x \in' j(V) = \{j(x) : x \in' V\}$ —$x \in' V$ iff $x = j(y)$ for some y;
- since any condition P in the language of equality and type-level membership is stratified, the set $\{x : P\}$ exists —thus, the set $\{x \in' V : P\}$ exists for each P— since each of its elements is an image under Endo, it has an inverse image S under Endo, such that $x \in' S$ iff $x \in \{x \in' V : P\}$ iff $x \in' V \wedge P$;
- if each "element" of x is an "element" of y (i. e., $j(x)$ is a subset of $j(y)$) and y belongs to V, it follows that y has an inverse image Y under j, and that the set of elements of Y which are inverse images under j of elements of x is an inverse image under j of x, so that $x = j(X)$ for some X and $x \in' V$. (A small note: the type-level membership can be used to define the Quine pair in terms of Endo).

The theory of type-level membership is axiomatized as follows. Its language is first-order logic with membership, equality, the constant V, and the function symbol j. Its non-logical axioms are:

(Ext) Objects with the same elements are equal.

(Iso) $j(x) \in j(y)$ iff $x \in y$; $j(x) = j(y)$ iff $x = y$; $x = j(y)$ for some y iff $x \in V$.

(Comp) For each condition P in which j is mentioned only in parameters, $\{x \in V : P\}$ exists (i. e., there is an object S such that $x \in S$ iff $x \in V \wedge P$).

(Subsets) Any subset of an element of V is an element of V.

This theory, which we will call B, for "bottomless set theory" (it is called B$-$ in [9], and the symbol B is used there for a related theory which turns out to be inconsistent), interprets NFU + Endomorphism as follows: define $x \in' y$ as $j(x) \in y \wedge y$ *is a subset of* V. Satisfaction of *(Ext)* and *(Nonset)* is immediate. We show that *(Comp)* is satisfied. We show first that $\{x : P\}$ exists for stratified conditions P in which x can be assigned type 0 and no variable assigned a positive type. Translate P into a formula P' in the language of B in the obvious way. By applying the automorphism j to both sides of the atomic formulæ of P', we can convert P' to a form in which each variable assigned type $-n$ appears with exactly n applications of j. Replace each $j^n(y)$ with the variable y restricted to $j^n(V)$. The resulting formula is a legitimate set definition in B and defines the set of x which satisfy P'; the image under j of this set has the desired members in the sense of the interpretation of NFU. Now observe that for any stratified condition P, the set of sufficiently iterated singletons of objects which satisfy P has the special form used above. It is sufficient to show that the union of any set of singletons exists in the interpretation of NFU: A *is a set of singletons* in the sense of NFU translates to A *is a set of singletons of elements of* $j(V)$ in B; the union of this set in the sense of B exists and is a subset of $j(V)$, an element of V, so by *(Subsets)* is an element of V and has an inverse image under j which is the union of this set in the sense of the interpreted NFU. Endo is interpreted by the map which takes singletons of elements of $j(V)$ to their elements.

The full axiom of extensionality in NFU (which would give NF) is interpreted in B as the assertion that all objects are subsets of V, or that all elements belong to V. We showed in [9] that the theory B with this additional assumption is inconsistent, even if we drop the axiom

(Subsets). Thus, the Axiom of Endomorphism is false in NF. It is interesting to note that the axioms of B are very similar to the first four axioms of the system W with negative types defined by Zimbarg and Hiller in [10]; their fifth axiom can thus be interpreted as defining a powerful extension of NFU + Endomorphism (W is considerably stronger than ZF).

We use a construction indicated by Forster in [11] to interpret NFU + Endomorphism in NFU + Infinity (it works in NFU, but some type differentials will be different). We refer to the union of the domain and range of a relation as its *full domain*. We call a relation R *well-founded* if every subset A of the full domain of R has an element a such that $b\ R\ a$ does not hold for any element of A. We call a relation R *extensional* if for every a and b in the full domain of R, $a \neq b$ implies that there is c such that $c\ R\ a$ holds and $c\ R\ b$ does not hold, or vice versa. We say that a well-founded extensional relation R *has a top node* if there is an element a of the full domain of R such that for each element b of the full domain, there is a finite sequence $\{b_i\}$ for $0 \leq i \leq n$ such that $b_0 = b$, $b_n = a$, and $b_i\ R\ b_{i+1}$ for each applicable i. A well-founded extensional relation with a top node can be understood as a picture of a set in the usual set theory. An *immediate subgraph* of a well-founded extensional relation with a top node a is associated with each node c such that $c\ R\ a$: the immediate subgraph is the largest well-founded extensional relation with top node c which is a subgraph of the original graph. The *singleton image* of a well-founded extensional graph with a top node is the graph which results if each node of the original graph is replaced with its singleton; since the unrestricted singleton map is not a function, the singleton image cannot be expected to be isomorphic to the original graph. We say that two well-founded extensional graphs with top node are *equivalent* if each immediate subgraph of each of them is equivalent to an immediate subgraph of the other; this inductive definition succeeds because of the well-foundedness of the graphs (start with graphs with only one node and no immediate subgraphs and work up). We say that a well-founded extensional graph R with top node is *saturated* if there is a well-ordered sequence of nodes of R, called *stages* of R, including the top node a of R, such that if s is a stage of R and B is a subset of the full domain of R such that for each $c \in B$ there is a stage t of R preceding s with $c\ R\ t$, then there is a node b such that $b\ R\ s$ and for each x in the full domain of R, $x\ R\ b$ exactly if $x \in B$. A saturated extensional well-founded graph with top node looks like a picture of an ordinal indexed power set of the empty set, a stage in the construction of the universe of the usual set theory. We call a well-founded extensional graph with top node a *picture* if it is a subgraph of some saturated well-founded extensional graph with top node; a saturated graph is a picture and will be referred to as a *saturated picture*. We take equivalence classes of pictures and define type level membership $\in_0$ by $[x] \in_0 [y]$ *iff x is equivalent to an immediate subgraph of y*. If we add a top node to the domain of $\in_0$, we obtain a relation which is easily seen to be a saturated picture. The equivalence class of this picture is found in the full domain of the extended $\in_0$, of course, and it is a stage of the extended $\in_0$, but it is not equivalent to the top node —the inductive argument which would appear to show this is unstratified. One can prove by induction that the maximal subgraph of $\in_0$ with top node the equivalence class of a given picture P is equivalent to the double singleton image of P. Since the double singleton image of $\in_0$ is not isomorphic to $\in_0$, neither is the singleton image; it is easy to see that the operation of taking singleton images of graphs induces an automorphism from the full domain of $\in_0$ into a proper subset of itself. In fact, the full domain of $\in_0$ with membership $\in_0$ and automorphism the singleton image interprets B, and so can interpret NFU + Endomorphism, if we define membership $[x] \in [y]$ as *the singleton image of x is equivalent to an immediate subgraph of y and all immediate subgraphs of y are equivalent to singleton images*. The restriction to saturated pictures, which is not necessary for simply interpreting

NFU + Endomorphism, has the effect of making the full domain of $\in_0$ look precisely like our model U; the full domain of $\in_0$, like U, interprets an ordinal indexed iterated power set of the empty set.

Observe that this technique for interpreting NFU + Endomorphism in NFU succeeds in NF, but yields an interpretation of NFU with many ur-elements. In [9], we presented a similar construction which yielded an interpretation of NFU + Endomorphism in NFU + Infinity which satisfies an analogue for NFU of the Anti-Foundation Axiom of Peter Aczel (see [12]); strictly speaking, the associated interpretation of B satisfies AFA.

Another fragment of NF which has been shown to be consistent is Marcel Crabbé's NFI, which we call *mildly impredicative* NF. NFI differs from NF in having a further restriction on the formula P in the axiom schema *(Comp)*: it must be possible to assign types to variables in P in such a way that no variable is assigned type higher than the type of the set being defined ($\{x : P\}$ exists if P can be stratified with x assigned type -1 and no variable assigned positive type). NFP, or *predicative* NF, has the further restriction that variables of the exact type of the set being defined cannot be bound. In [6], Crabbé showed that NFI interprets second-order arithmetic and can be proven to be consistent in third-order arithmetic. We show how to model NFI using a modified version of the technique used above to model NFU.

Observe that for any infinite set X, we can construct an ultrapower model of X with the same cardinality as the power set of X: let f be a bijection from X onto the set of finite subsets of X, and choose an ultrafilter $\mathcal{U}$ over X containing $\{x \in X : y \in f(x)\}$ for each $y \in X$. The ultrafilter $\mathcal{U}$ can be regarded as giving a description of a finite set which contains each standard element of X. The set of equivalence classes $[g]$ of functions g from X to X under the equivalence relation $g\ Q\ h$ defined as "$\{x \in X : g(x) = h(x)\} \in \mathcal{U}$" can be interpreted as a nonstandard model of X in the usual way; its cardinality is obviously no greater than that of the power set of X, but it is also at least as large as that of the power set of X, because the intersections of the finite set containing all standard elements of X encoded by $\mathcal{U}$ with the standard subsets of X are all distinct and all encoded by elements of the nonstandard model of X.

Take a transitive set X and define X_α for each ordinal α as above. Choose a $\emptyset_\beta$ such that $X_\alpha \in \emptyset_\beta$ for each $\alpha < \beta$, as before. For each $\alpha < \beta$, let f_α be a bijection from an ultrapower model of X_α to $X_{\alpha+1}$, sending each element of X_α to the corresponding standard object in the model of X_α. Note that the collection of f_α's can be used to interpret X_γ as an iterated ultrapower model of X_δ for each $\delta < \gamma$. Let U be a suitable nonstandard model of $\emptyset_\beta$ with an automorphism j moving an ordinal α to a smaller ordinal $j(\alpha)$.

We develop an interpretation of NFI. We begin by interpreting a type theory. Type 0 will be $X_{\alpha+1}$; type -1 will be X_α; membership of type -1 objects in type 0 objects will be the usual membership of U. Observe that type -1 can be interpreted as an iterated ultrapower model of $X_{\alpha'+1}$, the image of type 0 under the automorphism. Suppose that types $-n$ to 0 have been constructed, type $-n$ is a subset of type $-n-1$, and types $-n$ to -1 are interpreted as an iterated ultrapower model of the image under the automorphism of types $-n+1$ to 0. Since we know how to extend the structure with types $-n+1$ to 0 to include type $-n$, we know how to extend the structure with types $-n$ to -1 to include type $-n-1$ in such a way as to preserve the condition: the image under the automorphism of type $-n+1$ includes the image under the automorphism of type $-n$, and the membership relation of type $-n$ in type $-n-1$ has an image under the automorphism; these objects have corresponding extensions in the iterated ultrapower model, which we take as type $-n-1$ and membership of type $-n-1$ objects in type $-n$. It is obvious that the inductive hypothesis is preserved.

What kind of type theory is this? For any formula P in the language of the type theory and variable x of type -1, there is a set $\{x : P\}$ of type 0 (the lower types and their membership relations are present as objects in type 0). We certainly have extensionality for type 0 sets. Finally, we have typical ambiguity; the structure made up of types $-n$ to 0 is elementarily equivalent to the structure made up of types $-n-1$ to type -1 for every n, so decrementing the value of every type index in a formula will not affect its truth value. Applying typical ambiguity, we see that we have extensionality in every type, and comprehension in every type for formulæ in which no variable has type higher than that of the set being defined (the formula defining a set at type 0 will not contain any positive types, since there aren't any). We have the theory TTI of [6] with non-positive types and typical ambiguity, which is easily seen from Crabbé's remarks in [6] to be equiconsistent with NFI. We do not have the axiom of union, which would give us TT; at each step we take arbitrary subsets of a nonstandard model of the previous type, so an external set of objects interpreted as singletons will be an example of a set without a union.

Each type n of this model of TTI has a natural elementary embedding j_n into type $n-1$, sending each x to the "standard object" corresponding to its image under the automorphism. Consider the structure made up of sequences $\{s_i\}$ whose ith element is an element of type $-i$ and such that for each i, $s_{i+1} = j_i(s_i)$ or s_i is the empty set and s_{i+1} has no inverse image under j_i. Define "membership": $s \in' t$ iff $s_{i+1} \in t_i$ for all sufficiently large i. We claim that this structure is a model of NFI. Take any stratified condition P in the language of equality and membership on this structure; replace each variable of type $-n$ with a reference to its $k+n'$th term, with a suitable change in the membership relations used, and one has obtained a formula of TTI. If k is taken large enough that we have "significant" terms of each parameter in P, the truth value of the translated formula is independent of k, because of typical ambiguity of TTI, and easily seen to be the actual truth value of P. It follows directly that extensionality and comprehension of NFI hold in this structure, since the typed analogues of extensionality and each instance of comprehension hold in sufficiently low types of TTI.

An interesting feature of this model of NFI is that it turns out to interpret nth order arithmetic for every n; this shows that NFI can be strengthened considerably without paradox. In [9], we showed that the Axiom of Endomorphism is false in NFI; there is an encoding of the automorphism as a type-raising operation in these models, but its properties are more complex than they are in the case of NFU. The effects on the model of NFI of stronger assumptions about the size of X and the character of the automorphism of U are not as easy to determine as they are in the case of NFU.

We very briefly discuss the application of the technique to the construction of theories of functions of universal domain related to NF, of the type introduced in [13] and [14] by the author. The construction is based on a type hierarchy of function spaces rather than power sets: type 0 is an infinite set with a surjective pair; type $n+1$ is the set of all functions from type n to type n, with a surjective pair defined so that $(f, g)(x) = (f(x), g(x))$ for f, g of type

n and x of type $n-1$, with the pair on the left the pair already defined on type n. We need to modify this so that the types will be nested —otherwise it is entirely unclear how to define function spaces indexed by infinite ordinals. The simplest way to do this is as follows: provide an object *?* of type 0 such that $(?, ?) = ?$; define a_+ for each a of type 0 as the constant function of type 1 with value a everywhere; define a_+ for each a of type n as the function of type $n+1$ which takes b_+ to $a(b)_+$ for each b of type $n-1$ and takes each object of type n which is not b_+ for any b to *?*. Take the union of the types and identify a with a_+ for every a. Since function spaces are now nested, it is possible to define type α for α a limit ordinal as the union of types β for $\beta < \alpha$ and use the same definition for successor types as in the finite case. Suppose we have a nonstandard model of a type β with a type α taken to a lower type $j[\alpha]$ by an automorphism j. If we redefine $f(x)$ as $j^{-1}[f](x)$ for each f and x in type α, we obtain a model of the theory TRCL defined in [14]. A complete description of the construction of a model of the theory TRCU, an inessentially stronger version of TRCL defined in [13] and [14], is given in the author's thesis [13]. Without the formal definition of the theory or the proof, which is similar to the proofs given above, we remark that TRCL is an untyped lambda-calculus (with surjective pair identified with product and a characteristic function of equality) with a stratification restriction on abstraction analogous to the restriction on comprehension in NFU; it was shown in [14] that TRCL is precisely equivalent in consistency strength and expressive power to NFU + Infinity (this was actually shown for the more complicated theory TRCU, but it is obvious how to modify the proof). We could similarly model a lambda-calculus TRCI analogous to NFI.

References

[1] Willard van Orman Quine. New foundations for mathematical logic. *American Mathematical Monthly*, 44:70–80, 1937.

[2] Theodore Hailperin. A set of axioms for logic. *Journal of Symbolic Logic*, 9(1):1–19, March 1944.

[3] Ernst P. Specker. The axiom of choice in Quine's New Foundations for Mathematical Logic. *Proceed. Nat. Ac. Sc. U.S.A.*, 39:972–975, 1953.

[4] Maurice Boffa. The point on Quine's NF (with a bibliography). *Teoria*, IV/2:3–13, 1984.

[5] R.B. Jensen. On the consistency of a slight (?) modification of Quine's New Foundations. *Synthese*, 19:250–263, 1969.

[6] Marcel Crabbé. On the consistency of an impredicative subsystem of Quine's NF. *Journal of Symbolic Logic*, 47:131–136, 1982.

[7] Maurice Boffa. ZFJ and the consistency problem for NF. *Kurt Godel Gesellschaft Yearbook*, pages 102–106, 1988.

[8] Ernst P. Specker. Typical ambiguity. In Nagel, Suppes, and Tarski, editors, *Logic, Methodology and Philosophy of Science, Proceedings of the international Congress, Stanford, California, 1960*, pages 116–124, Stanford, California, 1962. Stanford University Press.

[9] M. Randall Holmes. The Axiom of Anti-Foundation in Jensen's "New Foundations with Ur-Elements". To appear.

[10] A.P. Hiller and J.P. Zimbarg. Self-reference with negative types. *Journal of Symbolic Logic*, 49(3):754–773, 1984.

[11] T.E. Forster. *Quine's New Foundations (an introduction)*. Number 5 in Cahiers du Centre de Logique. Cabay, Louvain-la-Neuve, 1983.

[12] Peter Aczel. *Non-well-founded sets*. CSLI. Stanford, 1988.

[13] M. Randall Holmes. *Systems of Combinatory Logic Related to Quine's "New Foundations"*. PhD thesis, State University of New York at Binghamton, 1990.

[14] M. Randall Holmes. Systems of Combinatory Logic Related to Quine's "New Foundations". To appear in APAL, 1991.