Annals of Pure and Applied Logic 53 (1991) 103–133
North-Holland

Systems of combinatory logic related to Quine's 'New Foundations'

M. Randall Holmes

Faculty of Science, University of Mons-Hainaut, 15 av. Maistriau, B7000 Mons, Belgium

Communicated by A. Nerode
Received 1 April 1990
Revised 16 December 1990

Abstract

Holmes, M.R., Systems of combinatory logic related to Quine's 'New Foundations', Annals of Pure and Applied Logic 53 (1991) 103–133.

Systems TRC and TRCU of illative combinatory logic are introduced and shown to be equivalent in consistency strength and expressive power to Quine's set theory 'New Foundations' (NF) and the fragment NFU + Infinity of NF described by Jensen, respectively. Jensen demonstrated the consistency of NFU + Infinity relative to ZFC; the question of the consistency of NF remains open. TRC and TRCU are presented here as classical first-order theories, although they can be presented as equational theories; they are not constructive.

Part 0. General introduction

In this paper, we work with two systems of untyped illative combinatory logic, first introduced in the author's thesis [9], one of which is provably consistent relative to the usual set theory and one of which is consistent if and only if Quine's set theory 'New Foundations' (introduced in [16]; the axioms of NF are stated in Part 1 below) is consistent. The consistency of 'New Foundations' remains an open problem ([2] is a discussion of this problem with a bibliography). We call these systems TRC (for 'type-respecting combinators') and TRCU (for 'TRC with ur-elements').

Atomic terms of TRC are Abst, Eq, p_1, p_2 and variables. If f and g are terms of TRC, $f(g)$, (f, g), and $K[f]$ are terms of TRC—these constructions are intended to represent function application, ordered pairing, and the formation of constant functions, respectively. We always write $f(g, h)$ instead of $f((g, h))$, and we write $nK[f]$ for the result of n applications of the constant function construction to f. We define Id as (p_1, p_2). The axioms of TRC are the axioms of

0168-0072/91/\$03.50

first-order logic with equality, which we do not state, and the following additional axioms:

I. $K[f](g) = f$.
II. $p_i(f_1, f_2) = f_i$ for $i = 1, 2$.
III. $(p_1(f), p_2(f)) = f$.
IV. $(f, g)(h) = (f(h), g(h))$.
V. $\mathrm{Abst}(f)(g)(h) = f(K[h])(g(h))$.
VI. $\mathrm{Eq}(f, g) = p_1$ if $f = g$; $\mathrm{Eq}(f, g) = p_2$ if $f \neq g$.
VII. If for all x, $f(x) = g(x)$, then $f = g$.
VIII. $p_1 \neq p_2$.

Note that it follows from Axioms III and IV that $\mathrm{Id}(f) = f$; Id is the identity function.

TRCU differs from TRC in the following ways: there is an additional atomic term ?, with an additional axiom $K[?] = (?, ?) = ?$, and Axiom VII, the Axiom of Extensionality of TRC, is replaced by two weaker axioms, in such a way as to allow the presence of 'ur-elements'. An ur-element' is defined as an object a such that $a(x) = ?$ for each x, but a is not equal to ?. To state the two axioms of extensionality of TRCU, we need to define the concept of 'iterated projection': the collection of iterated projections of an object x is the smallest collection which contains x and is closed under application of the projection operators p_1 and p_2. In Part 2, we show techniques which make it possible to encode this concept in the language of TRCU; the use of English in the statement of the axioms below is a matter of convenience. Here are the additional axioms of TRCU:

0. $K[?] = (?, ?) = ?$.
VIIA. If for all x, $f(x) = g(x)$, then $f = g$ or some iterated projection of f or g is an ur-element and distinct from the corresponding iterated projection of g or f.
VIIB. No iterated projection of $K[f]$, $\mathrm{Abst}(f)$, or $\mathrm{Abst}(f)(g)$ is an ur-element, for any f or g.

The work of Curry, Howard, and Girard (see [6]) has shown a close relationship between deductions in logics such as intuitionistic propositional logic or intuitionistic type theory and terms in related typed combinatory logics (our use of the term 'combinatory logic' includes the closely related field of lambda-calculus). Although our work here is not in the style of the work above, it may indicate the existence of similar connections between theories connected with Quine's 'New Foundations' and untyped systems of combinatory logic similar to our TRC (types correspond to propositions and can be thought of as tags on terms or equivalence classes of terms; in the untyped systems TRC(U), there is a notion of 'relative' type of a subterm relative to the term in which it is embedded, which would be expected to perform a similar function). We have begun investigating such connections; for instance, the Curry–Howard isomorphism, when applied to TRC, gives a distinctive system of propositional logic.

We will abbreviate 'New Foundations' as NF hereafter. We will also discuss Jensen's theory NFU (for 'NF with ur-elements') which he has shown to be consistent relative to the usual set theory (in [11]). We will show that TRC and NF are of the same consistency strength; each can interpret the other. Similarly, we will show that TRCU can interpret NFU with the Axiom of Infinity, and that NFU with the Axiom of Infinity can interpret TRCU. We actually interpret TRCU in a model of NFU whose elements are the 'sets of sets' in the original theory NFU + Infinity. The properties of this model enable us to present the interpretations of TRC in NF and TRCU in NFU in parallel. NFU with additional axioms such as the Axiom of Choice and strong mathematical induction (Jensen showed in [11] that NFU with these two axioms can be modelled) is equivalent to TRCU with similar axioms.

The original program of Curry for combinatory logic is described in the introduction to [5]. He listed four requirements: there should be no types or categories of objects; there should be an operation of application of a function to its argument; there should be an equality with the usual properties; the system should be 'combinatorially complete' — any function we can define intuitively by means of expressions containing variables can be represented in the system. Another basic aim is to show the possibility of eliminating the use of bound variables. We believe that TRCU basically fulfills this program, with a qualification. TRCU is a theory of functions with unrestricted application of functions to arguments and no typing. The notion of equality, with the usual properties, is present in two senses: TRCU is a first-order theory with equality, and it also includes a function which is essentially the characteristic function of the universal equality relation (this is what makes TRCU an 'illative' combinatory logic). The qualification is that TRCU is not quite combinatorially complete: there is a restriction on functional abstraction in TRC(U) analogous to the restriction of set comprehension in NF to 'stratified' conditions (Theorem 1 of this paper is the basic abstraction theorem for TRC(U)). Some qualification of the goals of Curry's program must be expected in a combinatory logic capable of expressing notions of propositional logic internally; a theory satisfying these goals exactly and containing a function corresponding to negation must be inconsistent. Because TRCU is a first-order theory, it is not possible to eliminate all uses of bound variables: in particular, the use of bound variables in quantification over 'unstratified' conditions cannot be eliminated. However, adding 'external functions' to the theory, analogous to the addition of 'proper classes' to ZF in Gödel–Bernays set theory, enables us to eliminate the use of bound variables. The use of 'proper classes' which cannot be elements adds expressive strength to ZF, making it much easier to state the Axiom of Replacement, for instance (Quine, in his theory ML of [18], added an impredicative theory of proper classes to 'New Foundations'). The theory with 'external functions' does not use bound variables, and is an equational theory rather than a first-order theory. The restriction on the 'external' functions is that they cannot be arguments or values of functions. The system is fully combinatorially complete, unlike TRCU (if one does not allow variables in names of external functions). However, it fails to completely satisfy another point of Curry's program — it is a two-sorted theory, with a distinction between 'internal' and 'external' functions. There is an equational version of TRC(U) in which all stratified theorems can be proven.

We believe that NFU and TRCU (but not NF or TRC) deserve consideration as alternative foundations for mathematics, when suitably extended; we also believe that TRCU may have some applications in theoretical computer science. We discuss these issues further in Part 5 below.

Part 1. Introduction to NF and combinatory logic

Introducing notation

We describe our notation for first-order logic. Primitive notions are taken to be "for all x, P", written "$(\forall x)(P)$", "not P", written "$\sim P$", and "P and Q", written "$P \,\&\, Q$". Other notions, considered to be defined in usual ways, are "P or Q", written "$P \vee Q$", "if P then Q", written "$P \rightarrow Q$", and "P if and only if Q", written "$P \leftrightarrow Q$". A defined quantifier is "for some x, P", written "$(\exists x)(P)$". The description operator "the x such that P" will be written "$(\mathrm{T}x)(P)$". The atomic formula "x is equal to y" will be written "$x = y$"; the atomic formula "x is an element of y" will be written "$x \in y$"; "$\sim x = y$" will be written "$x \neq y$". In type theory, quantifiers restricted to type n, "for all x of type n, P" and "for some x of type n, P", will be written "$(\forall x[n])(P)$" and "$(\exists x[n])(P)$".

We use the notation $T \models P$, where T is a theory (or a specific model of a theory) and P is a sentence to mean "the sentence P holds in (the given model of) T". When we define the notion of satisfaction in a model using this notation, we will usually only define it for atomic sentences; the definitions of satisfaction of complex sentences in terms of satisfaction of atomic sentences will be the natural ones.

Type theories

The first set theory we will describe here is the simple theory of types (abbreviated TT), TT has a sequence of universes of discourse, called 'types', indexed by the non-negative integers. Type 0 is a collection of individuals; type 1 is the collection of all collections of type 0 objects; type 2 is the collection of all collections of type 1 objects, and so forth. This can be expressed in three axioms: the Axiom of Extensionality asserts that if two objects of the same positive type have the same elements, they are the same object; the Axiom (scheme) of Comprehension asserts that for each formula P expressed in the language of first-order logic with equality, membership and type, the formula "$(\exists y[n+1])(\forall x[n])(x \in y \leftrightarrow P)$" is an axiom, where x is any variable, n is any non-negative integer, and y is any variable which does not occur in P; the Axiom (scheme) of Stratification asserts that each element of an object of type $n+1$ is of type n and that objects which belong to distinct types are distinct. Note that the restriction to positive types in the axiom of extensionality is essential if there is to be more than one object of type 0.

The simple theory of types is easily modelled in the usual set theory: let type 0 be represented by the cartesian product of any set X and $\{0\}$ and let type n be represented by the cartesian product of the nth iterated power set of X with $\{n\}$. Let $\mathrm{TT} \models (x, m) \in (y, n)$ be defined as "$x \in y \,\&\, n = m + 1$" in the sense of the usual set theory. It is easy to verify that the axioms given above for TT hold in the model.

Any proof in the simple theory of types remains a proof if all types mentioned are increased by a constant amount. Thus, it is possible to argue that the simple theory of types can be extended to a theory with types indexed by all the integers, and no base type, referred to as TNT, for 'theory of negative types'. Suppose that TNT were inconsistent; any proof of a contradiction would mention only finitely many types — raise the indices of the types so that all were non-negative, and one would produce a proof of a contradiction in the simple theory of types. TNT is more difficult to model: in a model of set theory with a nonstandard natural number N, build a model of the simple theory of types, and let type $N+n$ in the simple theory of types represent type n in TNT for each standard integer n.

Type theory used to motivate definition of NF

In TNT, any proof remains a proof if all types mentioned are increased or decreased by any constant amount. A natural question to raise is whether there is a model of TNT in which any true statement remains true if all types mentioned are increased or decreased by any constant amount — whether there is a model of TNT in which all the types are isomorphic to one another. This question is equivalent to the question of whether Quine's set theory NF is consistent. This approach to NF is taken from Specker's [15].

A one-to-one map h from a model of TNT onto itself is called a 'shifting automorphism' of the model if it sends each object of type n to an object of type $n+1$ and $x \in y$ iff $h(x) \in h(y)$ for all x and y in the model. A model of 'New Foundations' is constructed from a model of TNT with a shifting automorphism as follows: let the objects of the model of NF be the objects of type 0 in the model of TNT, and define $\mathrm{NF} \models x \in y$ as $\mathrm{TNT} \models x \in h(y)$ (recall that we use the notation $T \models P$ to represent "the sentence P is true in (a given model of) the theory T").

NF is an untyped theory; all types are collapsed into one. NF satisfies a simple Axiom of Extensionality — if NF $\models$ "x and y have the same elements", then TNT $\models$ "$h(x)$ and $h(y)$ have the same elements", from which it follows that $h(x) = h(y)$, since TNT $\models$ Extensionality, so $x = y$, since h is one-to-one.

We consider what kind of axiom schema of comprehension NF might satisfy. NF $\models P$, where P is a formula of first-order logic with equality and membership, is equivalent to TNT $\models P'$, where P' is a formula of first-order logic with equality, membership, and the shifting automorphism h, with all variables restricted to type 0. The axiom of comprehension of TNT will allow us to conclude that a formula of this type can be used to define a set if we can eliminate references to the shifting automorphism h, which does not appear in the usual language of TNT. We describe a procedure by which this can be done under some circumstances: since $x = y$ iff $h(x) = h(y)$ and $x \in y$ iff $h(x) \in h(y)$ in the model of TNT, we can apply h as many times as we wish to each side of each equation or atomic formula of membership that appears in our original formula. We can eliminate all applications of h to a variable if we can convert the formula into one in which the given variable appears everywhere with the same fixed number n of applications of h; we then replace each occurrence of the original variable (which is restricted to type 0) with n applications of h with a single variable with no applications of h which is restricted to type n. Since n-fold iteration of h yields a bijection between type 0 and type n, this gives an equivalent formula. Repeating this procedure for each variable, if possible, yields an equivalent formula in which h does not appear.

We can describe the conditions under which a formula in first-order logic with equality and membership can be translated from the language of NF into the language of TNT without mention of the shifting automorphism in this way: the procedure outlined above will work if it is possible to assign to each variable in the original formula an integer, referred to as a 'type', in such a way that each variable is assigned the same type wherever it appears, the types of x and y are the same in each subformula "$x = y$", and the type of y is one greater than the type of x in each subformula "$x \in y$". A formula satisfying this condition is said to be 'stratified'. A stratified formula can be translated into the language of TNT with h, then h can be applied to each side of each atomic formula until each variable with type n appears with n applications of h wherever it appears, and can be replaced with a new variable without applications of h restricted to type n wherever it appears. Suppose a stratified formula P contains the variable x with type n; if P' is the corresponding formula of TNT, there is a set A' of type $n+1$ such that "$(\forall x[n])(x \in A' \leftrightarrow P')$" holds. Let A be the set of type 0 such that applying h $n+1$ times to A gives A'. It follows that TNT $\models (\forall x[0])(x \in h(A) \leftrightarrow P')$. But this is equivalent to the assertion NF $\models (\forall x)(x \in A \leftrightarrow P)$. The Axiom of Comprehension of NF asserts that for each stratified formula P, the formula "$(\exists y)(\forall x)(x \in y \leftrightarrow P)$" holds for each variable x and each variable y which does not appear in P.

We summarize the definition of NF. NF is a first-order theory with equality and membership. A formula of NF is said to be 'stratified' if integer 'types' can be assigned to the variables appearing in it in such a way that in each subformula "$x = y$" the types of x and y are the same, and in each subformula "$x \in y$" the type of y exceeds the type of x by one. NF has two non-logical axioms:

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

(Comp) $(\exists y)(\forall x)(x \in y \leftrightarrow P)$, for each stratified P, y not free in P.

Any model of the Axioms of Extensionality and Comprehension of NF can be converted into a model of TNT with a shifting automorphism: objects of type n in the resulting model of TNT will be ordered pairs (x, n), where x is an object of the model of NF, and the shifting automorphism h will take each pair (x, n) to $(x, n + 1)$. TNT $\vDash (x, m) \in (y, n)$ will translate to "NF $\vDash x \in y$ and $m + 1 = n$". It is straightforward to verify that this will actually yield a model of TNT with a shifting automorphism. Thus, the models of TNT with shifting automorphisms correspond exactly to the models of NF, the theory determined by the axioms of extensionality and comprehension given above.

The question of consistency: NFU introduced and compared with NF

The question of the consistency of NF, or, equivalently, of the existence of models of TNT with shifting automorphisms, remains open. The difficulty does not lie with the powerful Axiom of Comprehension of NF, where one might expect it to lie; the problem is with extensionality. The theory NFU, 'New foundations with ur-elements', obtained by weakening the axiom of extensionality to apply only to objects which have elements, is known to be consistent relative to the usual set theory. In NFU, we have many 'ur-elements' which have no elements but are distinct from one another. In all known models of NFU, most objects in the universe are 'ur-elements' — if it were possible to construct a model of NFU in which the cardinality of the set of ur-elements was less than or equal to the cardinality of the set of sets internally to the model, it would be possible to construct a model of NF.

Models of NFU correspond to models of a type theory with shifting automorphisms. The type theory in question is TNTU, the theory of negative types with ur-elements, in which the axiom of extensionality is weakened just as it is in NFU — for each n, type $n + 1$ contains all sets of elements of type n and a number of ur-elements. We indicate the construction of a model of NFU. Take a model of ZF (or a suitable fragment) on which there is an automorphism which sends a natural number N to a smaller number N'. The model of NFU is the Nth power set of a transitive set X. The automorphism induces an isomorphism between the Nth power set of X and the N'th power set of X. Define NFU $\vDash x \in y$ as "$x' \in y$ and y belongs to the $(N' + 1)$st power set of X", where x' is the image of x under the induced isomorphism. Note that if y is not in the $(N'+1)$st power set of X, y is interpreted as an ur-element. It is straightforward to verify that this is a model of NFU — in verifying the Axiom of Comprehension, reference to the isomorphism, which is not definable internally to the model of set theory, is eliminated much as reference to the shifting automorphism above is eliminated, by changing variables with n applications of the isomorphism to variables restricted to the image of the Nth power set of X under n applications of the isomorphism, an internally definable object. Maurice Boffa has used models of ZF with automorphisms to construct models of NFU in [3].

NF is a stronger theory than NFU. In NF, it is possible to disprove the Axiom of Choice in its most general form, and thus to prove the Axiom of Infinity. The axioms of NFU are compatible with the cardinality of the universe being finite (internally — any model of NFU is actually infinite and contains objects corresponding to all the standard natural numbers). NFU is consistent with the Axiom of Choice, and can in fact be extended with additional axioms to a theory on which all of mathematics could be founded. The same is probably true of NF if it is consistent — restricted forms of the Axiom of Choice adequate for all usual applications are probably consistent with NF.

Constructions and set-theoretical notation in NF

We carry out some constructions in NF, usually following Rosser's [12]. Some of these constructions will not be valid for NFU — in particular, the definition of the ordered pair given below, which is due to Quine (in [17]), will not work in NFU, because it relies on extensionality. We will show below how to get a similar definition of the ordered pair in an extension of NFU.

The language of NF is first-order logic with equality and membership. It is possible to augment this language with proper names of the form "the x such that P", where P is a formula, in the usual way due to Russell; we write this "$(\mathrm{T}x)(P)$". For purposes of stratification, the term "$(\mathrm{T}x)(P)$" is assigned the same type as the variable x. A proper name with no parameters can appear with different types in a stratified formula: change the bound variables in each occurrence to preserve stratification. If it is not the case that P holds for one and only one x, we consider "the x such that P" to be a name for the empty set. The introduction of proper names into NF is discussed by Rosser in [12].

The (possibly parametrized) proper name $\{x \mid P\}$ abbreviates "$(\mathrm{T}y)(\forall x)(x \in y \leftrightarrow P)$", where y is a variable which does not occur in P. If P is stratified. $\{x \mid P\}$ represents "the set of x such that P" — if P is not stratified, it may represent "the set of x such that P" or the empty set, depending on whether "the set of x such that P" exists. Some sets which do exist are $V = \{x \mid x = x\}$, the universe, and $\{\,\} = \{x \mid x \neq x\}$, the empty set. Parameterized proper names which can be defined include $-A = \{x \mid {\sim} x \in A\}$, $A + B = \{x \mid x \in A \vee x \in B\}$ and $AB = \{x \mid x \in A \mathbin{\&} x \in B\}$. In these names, the variables A and B appear with the same type as the whole proper name. In the parameterized proper name $\{a\} = \{x \mid x = a\}$, the type of the proper name is one greater than the type of the parameter a. We define $\{a, b\}$ as $\{a\} + \{b\}$.

Natural numbers have natural definitions in NF, similar to the definitions of natural numbers used in naive set theory before the discovery of the paradoxes. $0 = \{\{\ \}\}$, the set of all sets with zero elements. If n is defined and means "the set of all sets with n elements", the parameterized proper name

$$n + 1 = \{x \mid (\exists y)(\exists z)(y \in n \;\&\; z \in -y \;\&\; y + \{z\} = x\}$$

is stratified, has the same type as the parameter n, and clearly defines "the set of all sets with $n + 1$ elements". The set $\mathbb{N}$ of natural numbers is defined as

$$\{x \mid (\forall a)((0 \in a \;\&\; (\forall y)(y \in a \rightarrow y + 1 \in a)) \rightarrow x \in a)\}.$$

It is a difficult theorem of NF that $\mathbb{N}$ is infinite, which is equivalent to the assertion that $\{\ \} \in -\mathbb{N}$ (if the universe had finite cardinality n, then $n + 1$ would be the empty set). Notice that we distinguish between $n + 1$ and $n\ +\ 1$, the union of n and the set 1 of all one-element sets, by the use of spaces.

We can use the infinite set $\mathbb{N}$ to define ordered pairs. Everything we have done up to this point could be done in NFU, but this definition of the ordered pair, due to Quine in [17], relies on strong extensionality. The usual definition of the ordered pair as $\{\{x\}, \{x, y\}\}$ would be unsatisfactory in NF, because the type of (x, y) would not be the same as the types of its parameters. For a set A, we define $S[A]$, the image of A under the successor operation, as

$$\{x \mid x \in (-\mathbb{N})A \vee (\exists y)(y \in A \;\&\; x = y + 1)\}.$$

The parameterized proper name $S[A]$ is stratified and has the same type as its parameter. Now we define (A, B) as

$$\{x \mid (\exists y)(y \in A \;\&\; x = S[y]) \vee (\exists z)(z \in B \;\&\; x = S[z] + \{0\})).$$

It is straightforward to verify that (A, B) has the same type as its parameters and to define proper names $p_1(A)$ and $p_2(A)$ such that $p_1(A, B) = A$, $p_2(A, B) = B$, and $(p_1(A), p_2(A)) = A$. We have defined a surjective pairing relation.

Cardinal and ordinal numbers; the paradoxes avoided

Now that we have ordered pairs, we can proceed to develop the theory of functions and relations. In NFU, since it is possible for the universe to be finite, we cannot define a type-level pairing relation without additonal assumptions. In NFU with the Axiom of Infinity and the Axiom of Choice, we could prove the existence of a surjective pairing relation on the universe. In TRC(U), we will have pairing as an independent primitive. We will show in Part 3 how to model NFU + Infinity + "there is a surjective pairing relation" in NFU + Infinity.

Relations and functions are defined in the usual way in NF, with the odd feature that since every object is a pair, every set is a relation. (Every object is a pair because the Quine pair is surjective; it satisfies the equation

$(p_1(A), p_2(A)) = A$). Two sets are said to be 'equipotent' if there is a bijection which takes one onto the other. A cardinal number is an equivalence class under equipotence. $|A| = \{x \mid x \text{ is equipotent with } A\}$, the cardinality of A, is a stratified proper name with type one higher than that of its parameter, when written out completely. The Cantor theorem that the cardinality of the collection of all subsets of A is greater than the cardinality of A turns out not to hold in NF; what can be proven is that the cardinality of the set of all subsets of A is greater than the cardinality of the set of all one-element subsets of A. The intuitive one-to-one correspondence between the set A and the set of its one-element subsets is defined by an unstratified formula and cannot be realized by a bijection in NF. The Cantor paradox, the proof in naive set theory that the power set of the universe is larger than the universe, becomes in NF a proof that V, the universe (the set of all subsets of the universe) is larger than 1 (the set of all one-element subsets of the universe). This is counterintuitive but not paradoxical. In NFU + AC, the situation is even stranger. The cardinality of 1 is less than the cardinality of the set of all sets (the power set of the universe), which is in turn much smaller than the cardinality of the universe (which contains a large collection of ur-elements in addition to the sets).

We have shown how the Cantor paradox is avoided in NF(U). The Russell paradox is obviously avoided — the formula "$\sim x \in x$" which defines the Russell class is clearly not stratified. The Burali-Forti paradox of the largest ordinal is avoided in a stranger way. Ordinal numbers are defined in NF as equivalence classes of well-orderings — the usual definition of the ordinals due to von Neumann is unstratified. The collection of all ordinals has a natural well-ordering on it; the argument for the paradox considers the order-type of this well-ordering and concludes that it is larger than any ordinal. This turns out not to be the case in NF; the order type of the collection of all ordinals under the usual well-ordering turns out to be smaller than the largest ordinals. The natural correspondence between ordinals and the order-types of initial segments of the ordinals, which one would normally use to show that the order-type of the ordinals is larger than any ordinal, is defined by an unstratified condition and fails to exist in NF; an ordinal does not necessarily have the same order-type as the corresponding initial segment of the ordinals. It follows that any model of NF(U) in the usual set theory is nonstandard in the sense that the ordinals of the model will not actually be well-ordered: the sequence consisting of the order-type of all the ordinals, the order-type of the ordinals up to that order-type, and so on, with each term of the sequence the order-type of the ordinals less than the previous term, will be an infinite descending sequence. Of course, this sequence will not be definable inside NF (if NF is consistent) and cannot be defined in NFU, which is consistent.

Strongly Cantorian sets and the Axiom of Counting

The Cantor theorem holds for any set A such that there is a bijection between A and the set of one-element subsets of A. Such sets are called 'Cantorian'. The property 'Cantorian' is defined by an unstratified condition, and in fact cannot determine a set. A stronger property is the existence of a one-to-one correspondence between the set A and the collection of one-element subsets of A in which each element of A is associated with its own singleton. Sets with this property are called 'strongly Cantorian'. Strongly Cantorian sets are very well-behaved — we can think of them as the 'small' sets to which the usual set theory restricts itself. Any formula in which each variable is restricted to some strongly Cantorian set is equivalent to a stratified formula (obtained by using the correspondences between elements and singletons to raise and lower types), and thus defines a set. Products of strongly Cantorian sets and power sets of strongly Cantorian sets are strongly Cantorian. Any finite set actually defined by enumeration is strongly Cantorian, but it turns out that it is not possible to prove that all finite sets are strongly Cantorian or that $\mathbb{N}$ is strongly Cantorian without additional assumptions. Rosser added as an axiom in his [12] an assertion equivalent to the assertion "$\mathbb{N}$ is strongly Cantorian", which he called the Axiom of Counting. The Axiom of Counting or something stronger is required to make NF (or NFU) a reasonable system. It is possible to construct models of NFU in which all natural numbers are standard, so the Axiom of Counting and strong mathematical induction hold. NF has problems with induction because $\mathbb{N}$ is defined as the intersection of all inductive collections which are sets of NF — it does not follow that induction holds for conditions which do not define sets. It has been proven that the Axiom of Counting is independent of the other axioms of NF(U) (it does not hold, for instance, in the model of NFU described above). Rosser showed in [12] that NF with the Axiom of Counting and the Axiom of Denumerable Choice is a system strong enough to found most of classical mathematics (if it is consistent), and his development can be duplicated for NFU with the Axioms of Choice and Counting, which is known to be consistent (the only real problem is the lack of the Quine ordered pair; the existence of a type-level pairing relation can be proven using AC or one can show the relative consistency of the existence of a type-level pairing relation analogous to the Quine pair by constructing a model as we do below (this requires only Infinity); the fact that full Choice holds is a positive advantage over NF). In fact, axioms of infinity can be added to extend NFU so that it can interpret set theories as strong as desired.

For any infinite set X, it is possible to construct a nonstandard model of set theory with a proper automorphism which fixes X (X may have nonstandard elements in the model, but none of them are moved by the automorphism — see [9, pp. 62–64] (we do not claim originality!)). One can use this fact to build a model of NFU which satisfies the Axiom of Counting, or even stronger axioms asserting the existence of large strongly Cantorian sets. Let X be a transitive infinite set, and find a nonstandard model of ZF or a suitable fragment with an automorphism which fixes X. Let a be an ordinal which is moved by the automorphism to a smaller ordinal a'. The model of NFU will be the ath iterated power set of X (taking unions at limit ordinals). Define NFU $\vDash x \in y$, as above, as "$x' \in y$ and y belongs to the $(a' + 1)$st iterated power set of X". This is a model of

NFU for the same reasons that the first model described is a model of NFU, and it follows additionally, since elements of X are not moved by the automorphism, that the map from elements of X are not moved by the automorphism, tht the map from elements of X to their singletons is interpreted normally, so X is strongly Cantorian. If X is taken to be countable, this gives us the Axiom of Counting. If X is taken to be a model of ZFC (assuming, of course, the existence of an inaccessible cardinal), we have a model of NFU with a strongly Cantorian model of ZFC contained in it. Boffa has similar constructions in [3]. The Axiom of Choice also holds in all the models of NFU that we have described.

The Axiom of Choice fails in NF

Specker has shown in [14] that the Universe cannot be well-ordered in NF, so the Axiom of Choice is false. We give an argument similar to his for NF + Axiom of Counting. The usual exponentiation operation on cardinals is not a function in NF, because it is defined in an unstratified way. However, we can define a function $\exp(A)$, where A is a cardinal number, as "the cardinality of the power set of a set B such that B has the same cardinality as the set of one-element subsets of a set of cardinality A, or the empty set if there is no such B". $\exp(|V|)$ is the empty set, for instance. We can then define the set of cardinals such that a finite number of iterated applications of exp to the cardinal yield $|V|$, the cardinality of the universe. If the universe can be well-ordered, there is a smallest element of this set of cardinals. Consider the finite set of cardinals obtained by iterated application of exp to this smallest element. Let this set be represented as $\{A_0, \ldots, A_n\}$, where A_0 is the smallest element of the class defined above, $A_n = |V|$, and $A_{i+1} = \exp(A_i)$ for each i. For each cardinal A, let $[A]$ represent the cardinality of the set of one-element subsets of a set of cardinality A. $[|V|] = |1|$, for instance. The operation of constructing $[A]$ from A is unstratified, but the set $\{A_0, \ldots, A_n\}$ is finite, thus strongly Cantorian by the Axiom of Counting. It follows that the natural one-to-one correspondence between $\{A_0, \ldots, A_n\}$ and $\{[A_0], \ldots, [A_n]\}$ is definable in NF with the Axiom of Counting, using the fact that finite sets are strongly Cantorian to get around stratification restrictions (this is not necessarily true in NF, and Specker's argument becomes more complex at this point). It is straightforward to show that $\exp([A]) = [\exp(A)]$ for all A, so, by the definition of the class of which A_0 is the smallest element, $[A_0] \geqslant A_0$ — note that a finite number of applications of exp takes $[A_0]$ to $[A_n] = [|V|] = |1|$, and one more will take it to $|V|$. By induction, it follows that $[A_i] \geqslant A_i$ for each i (again, the Axiom of Counting is needed to raise or lowe types so that the induction can be carried out). But then $|1| = [A_n] \geqslant A_n = |V|$, a contradiction. The argument fails in NFU because $\exp(|1|) < |V|$ in NFU.

Since the Axiom of Choice is false in NF, the Axiom of Infinity must be true — if the universe cannot be well-ordered, it certainly cannot be finite! We have not justified this by the argument above; our additional assumption, the Axiom of Counting, implies the Axiom of Infinity by itself. NFU differs radically from NF in these respects; there are models of NFU in which the universe is finite, and all natural models of NFU that we have constructed actually satisfy the Axiom of Choice.

Combinatory Logic introduced

We describe a classical system of untyped combinatory logic (see [5] for combinatory logic and lambda-calculus). Objects in combinatory logic may be understood as functions of universal domain. Atomic terms in the system we describe here are I, K, S, and variables. If A and B are terms, $A(B)$ is a term, read as the application of A to B. Note that we write function application in the usual way, rather than as (AB), as is traditional in combinatory logic. The axioms, in addition to the usual axioms and rules for an equational system, are the following:

(I) $I(A) = A$.

(K) $K(A)(B) = A$.

(S) $S(A)(B)(C) = A(C)(B(C))$.

(E) If $A(x) = B(x)$, where x is a variable which does not occur in A or B, then $A = B$.

This system of combinatory logic has a powerful abstraction property. If we use the notation $P[A/x]$ to represent the result of substituting the term A for the variable x wherever it occurs in the term P, we can state it as follows: for each term P and variable x, there is a term $(\mathsf{L}x)(P)$ not containing x such that $(\mathsf{L}x)(P)(A) = P[A/x]$ for each term A. We define $(\mathsf{L}x)(P)$ by induction: $(\mathsf{L}x)(x) = I$; $(\mathsf{L}x)(a) = K(a)$, where a is an atomic term other than x; $(\mathsf{L}x)(A(B)) = S((\mathsf{L}x)(A))((\mathsf{L}x)(B))$ for any terms A and B. It is straightforward to verify by induction on the structure of terms that $(\mathsf{L}x)(P)$ has the desired properties. I can actually be defined in terms of S and K as $S(K)(K)$.

The abstraction property of combinatory logic is actually too powerful: it is not possible to adjoin simple notions of propositional logic to combinatory logic without inconsistency. Suppose Neg represented the negation operation: consider $R = (\mathsf{L}x)(\mathrm{Neg}(x(x)))$. Then $R(R) = \mathrm{Neg}(R(R))$ by the abstraction theorem. But negation cannot have a fixed point. Notice that we have essentially duplicated Russell's paradox. Notice also that the argument we have given could be used to show that any function in combinatory logic has a fixed point.

Combinatory logic can be proven to be consistent, both by syntactical methods and by the construction of models. It is in some sense the naive theory of functions, analogous to the original naive set theory which was shown to be false by the paradoxes. Since it is actually consistent, it fares better than the naive theory of sets, but it remains consistent by avoiding logical notions. Systems of combinatory logic which attempt to incorporate logical notions have been referred to as systems of 'illative' combinatory logic. They involve complicated restrictions on the use of logical notions. Systems of typed combinatory logic which accommodate logical notions very well have also been defined, but these are no longer theories of functions of universal domain — they are theories of functions of restricted domain, and the simplest systems of typed combinatory logic are easily modelled in set theory.

We will show that NF is equivalent to a system TRC of untyped illative combinatory logic, in the sense that sentences of each theory can be reinterpreted as sentences of the other in such a way that all theorems remain true. We will exhibit a theory TRCU, a modification of TRC to allow 'ur-elements' (a concept which has a slightly different interpretation in a theory of functions of universal domain), for which we can construct a model in the usual set theory (we will actually interpret TRCU in NFU, which is known to be consistent relative to the usual set theory).

Part 2. TRC is introduced and used to interpret NF

Just as we introduced a typed conventional set theory before introducing NF, we will introduce a typed theory of functions, which we will call 'typed TRC', in order to motivate the definition of TRC.

A model of 'typed TRC' consists of types indexed by the non-negative integers, like a model of the simple theory of types. Type 0 is a set X with a surjective pairing relation, a bijection f from the cartesian product of X with itself to X. Where x and y are elements of X, we will suppress mention of f and represent $f(x, y)$ by (x, y). We will have no further occasion to refer to the usual pairing operation; we will define additional pairing operations on the other types of typed TRC.

When type n has been defined, and the surjective pairing operation on type n has been defined, we define type $n+1$ as the set of functions from type n to type n, and we define (f, g), for f and g of type $n+1$, as the function which takes each h in type n to $(f(h), g(h))$, where the pairing used here is the pairing in type n already defined. It is straightforward to show that the pairing on type $n+1$ is also surjective.

We define some constants in type n for each sufficiently large n. In type $n+1$, we can define projection operators p_1 and p_2 such that $p_i(x_1, x_2) = x_i$ for $i = 1, 2$ and x_i of type n, also satisfying $(p_1(x), p_2(x)) = x$ for each x of type n, since the pairing operation on each type is surjective. In type $n+2$, we can define a constant Eq such that $\text{Eq}(x, y) = p_1$ if $x = y$ and p_2 if $x \neq y$, the constants p_1 and p_2 being those of type $n+1$, for each x and y of type $n+1$. In type $n+1$, we can define, for each a of type n, a function $K[a]$ such that $K[a](x) = a$ for each x of type n. In type $n+3$, we can define a constant Abst such that $\mathrm{Abst}(f)(g)(h) = f(K[h])(g(h))$ for each f, g, and h of types $n+2$, $n+1$, and n, respectively.

We give the formal description of the theory 'typed TRC'. Atomic terms of typed TRC consist of constants p_1 and p_2 of each type $n+1$, constants Eq for each type $n+2$, constants Abst of each type $n+3$, and variables of each type (all atoms actually have type superscripts, which we are suppressing). If f and g are terms of typed TRC of the same type n, (f, g) is a term of type n. If f is a term of type $n+1$ and g is a term of type n, $f(g)$ is a term of type n. We write $f(g, h)$ instead of $f((g, h))$. If f is a term of type n, $K[f]$ is a term of type $n+1$. Note that the type of any subterm of a term can be deduced from the type and structure of the term; this is why it is safe to avoid using type indices on terms. The axioms of typed TRC are the following, where the terms on each side of each axiom have the same type, which ranges over all types of typed TRC in which they make sense, in addition to axioms of first-order logic with equality. These axioms are consistent because they hold in the model described above.

I. $K[f](g) = f$.
II. $p_i(f_1, f_2) = f_i$ for $i = 1, 2$.
III. $(p_1(f), p_2(f)) = f$.
IV. $(f, g)(h) = (f(h), g(h))$.
V. $\mathrm{Abst}(f)(g)(h) = f(K[h])(g(h))$.
VI. $\mathrm{Eq}(f, g) = p_1$ if $f = g$; $\mathrm{Eq}(f, g) = p_2$ if $f \neq g$.
VII. If for all x, $f(x) = g(x)$, then $f = g$.
VIII. $p_1 \neq p_2$.

The way in which the axioms are presented makes it clear that a proof in typed TRC remains a proof if all types appearing are raised by the same constant amount. This in turn implies that we can allow all integer types in typed TRC without fear of contradiction. This is shown in the same way as it was shown for the simple theory of types above. We will refer to the expanded theory as 'typed TRC with integer types'. In typed TRC with integer types, all types contain the same kinds of terms, and any proof in typed TRC with integer types remains a proof if the same integer is added to all types appearing. We raise the same question: Is there a model of typed TRC with integer types in which all true statements remain true if the same integer is added to all types appearing? This is the same question as that of the consistency of the theory TRC which we are about to introduce, and it turns out to be exactly equivalent to the question of the consistency of 'New Foundations'.

A bijection H from a model of typed TRC with integer types to itself is said to be a 'shifting automorphism' if it takes each object of type n to an object of type $n+1$ and if $H[(f, g)] = (H[f], H[g])$ and $H[f(g)] = H[f](H[g])$ for each f and g in the model. One can prove that H respects the construction of constant functions from the fact that it respects application and the axiom of extensionality.

Given a model of typed TRC with integer types with a shifting automorphism H with inverse J, we construct a model of TRC. The objects of the model of TRC are the objects of type 0 in the model of typed TRC with integer types. "(f, g)" in the sense of TRC is defined as (f, g); "$f(g)$" in the sense of TRC is defined as $H[f](g)$; "$K[f]$" in the sense of TRC is defined as $J[K[f]]$. It is straightforward to verify that axioms I–VIII above with all reference to types suppressed hold in the model of TRC with this language.

TRC is defined as follows: atomic terms of TRC are p_1, p_2, Abst, Eq, and variables; if f and g are terms of TRC, (f, g), $f(g)$, and $K[f]$ are terms of TRC (we will always write $f(g, h)$ instead of $f((g, h))$, and $nK[f]$ to represent the result of n applications of K to f); the axioms of TRC are I–VIII above plus the axioms and rules of first-order logic with equality. Another extension of our notation follows from the observation that $(p_1, p_2)(x) = (p_1(x), p_2(x)) = x$ for all x; (p_1, p_2) is an identity function, and we define Id as (p_1, p_2). It is straightforward to construct a model of typed TRC with integer types with a shifting automorphism given a model of TRC, in essentially the same way as one constructs a model of TNT with a shifting automorphism given a model of NF: represent objects of type n by ordered pairs (f, n), where f is a term of TRC, and define all constructions in the obvious way, observing the type restrictions of typed TRC with integer types.

What is not clear here is the motivation for the choice of functions selected as constants of typed TRC initially. The following theorem (stated as a theorem about TRC, but valid in either of the precursor theories with suitable modifications) should indicate some of our motivation.

Before we can state the Abstraction Theorem of TRC, we have to recover the ability to talk about 'types' in TRC. Objects of TRC are untyped; we have the ability to type subterms of a term of TRC relative to the term. The type of a subterm of a term of TRC is defined as the type the subterm would have if the entire term were to be interpreted as a term of typed TRC with integer types of type 0. Equivalently, the type of a term relative to itself is 0; if the type of a subterm (f, g) is n, the types of f and g are n; if the type of $f(g)$ is n, the types of f and g are $n+1$ and n, respectively; if the type of $K[f]$ is n, the type of f is $n-1$. Note that what we are typing here are occurrences of subterms; the same term may appear as a subterm of a given term with different relative types. It is straightforward to demonstrate that if f occurs with type m in g, and g occurs with type n in h, then f occurs with type $m+n$ in h. We can now state the Abstraction Theorem for TRC (compare the abstraction theorem for combinatory logic above and the Axiom of Comprehension of NF).

Theorem 1. *Let T be a term of* TRC, *and let x be a variable which occurs in T with no type other than* 0. *Then there is a term $(\text{L}x)(T)$ which does not contain x*

such that "$(\forall x)((\mathrm{L}x)(T)(x) = T)$" is a theorem. Moreover, for each variable y other than x (which does not occur in $(\mathrm{L}x)(T)$), y occurs in $(\mathrm{L}x)(T)$ with type n exactly if y occurs in T with type $n + 1$.

Proof. We may assume without loss of generality that all subterms of T of the form $K[A]$ are of the form $nK[\text{atom}]$; we can use the theorems $K[(f, g)] = (K[f], K[g])$ and $K[f(g)] = \text{Abst}(K[f])(K[g])$, both easy consequences of extensionality, to eliminate all constant function expressions with complex arguments, observing that applications of these theorems leave types of variables unchanged.

We then define $(\mathrm{L}x)(T)$ by induction. $(\mathrm{L}x)(x) = \text{Id} = (p_1, p_2)$. $(\mathrm{L}x)(A) = K[A]$ when A is a term which does not contain the variable x. $(\mathrm{L}x)(U, V) = ((\mathrm{L}x)(U), (\mathrm{L}x)(V))$. $(\mathrm{L}x)(nK[\text{atom}]) = [n+1]K[\text{atom}]$, because the typing restrictions on x ensure that the atom is not x, so if $T = nK[\text{atom}]$, T does not contain x. The verifications of these cases present no difficulties.

The only case in an argument by induction on the length of terms that is not taken care of is the case $(\mathrm{L}x)(U(V))$. By the conditions on the type of x, x occurs in U with no type other than -1, which implies that it does not occur except in the context $K[x]$, because a subterm cannot appear with lower type than a subterm which contains it unless it is in the argument of a constant function expression, and all constant function expressions in U are of the form $nK[x]$ if they contain x. The expression $K[x]$ will occur in U with no type other than 0. Thus, the expression $U[x/K[x]]$ contains x with no type other than 0. We define $(\mathrm{L}x)(U(V))$ as $\text{Abst}((\mathrm{L}x)(U[x/K[x]]))((\mathrm{L}x)(V))$. We verify that $(\mathrm{L}x)(U(V))(x) = U(V)$:

$$\begin{aligned}&\text{Abst}((\mathrm{L}x)(U[x/K[x]]))((\mathrm{L}x)(V))(x)\\&\quad = (\mathrm{L}x)(U[x/K[x]])(K[x])((\mathrm{L}x)(V)(x))\\&\quad = U[x/K[x]][K[x]/x](V) = U(V).\end{aligned}$$

The assertions about types of variables follow in this case (as in the other cases, which are easy to verify) from the fact that types of variables are not changed in instances of application of axioms I–V, so, in general, the types with which any variable occurs in $(\mathrm{L}x)(T)(x)$ will be the same as the types with which it occurs in T, so the types with which it occurs in $(\mathrm{L}x)(T)$ will be one less. The theorem follows by induction on the length of terms. The proof of Theorem 1 is complete. □

Note that two consequences of the Axiom of Extensionality, $K[(f, g)] = (K[f], K[g])$ and $K[f(g)] = \text{Abst}(K[f])(K[g])$, were needed to prove Theorem 1. We have had occasion to consider weaker forms of TRC with a weaker axiom of extensionality (such as TRCU) or none (in [9]); it will be necessary in such theories either to verify that the weaker forms of extensionality imply these two theorems or, in the case of a weak TRC with no extensionality, to adjoin these two assertions as independent axioms. We are not interested in any form of 'weak TRC' in which Theorem 1 or something equivalent cannot be proven. The following Corollary to Theorem 1 improves our abstraction notation.

Corollary to Theorem 1. *Let T be a term of* TRC, *and let x and y be variables which do not occur in T with any type other than* 0. *Then there is a term $(\mathsf{L}xy)(T)$ such that "$(\forall x)(\forall y)((\mathsf{L}xy)(T)(x, y) = T)$" is a theorem, satisfying conditions on types of variables analogous to those in Theorem* 1.

Proof. Let z be a variable which does not occur in T. Let

$$(\mathsf{L}xy)(T) = (\mathsf{L}z)(T[p_1(z)/x, p_2(z)/y]).$$

Apply Theorem 1. The proof of the Corollary is complete. □

A remark on notation: in 'classical' combinatory logic or lambda-calculus, pairing does not appear as an independent notion. A trick called 'currying' is used instead, in which what in our notation would be written $f(a)(b)$, and in classical notation would be written fab, is used to represent the application of a function f to two arguments a and b; the pair (a, b) is represented by $(\mathsf{L}x)(x(a)(b))$. This is inappropriate in TRC, because the relative types of the two arguments of f or terms of the pair would be different. The fact that this method of handling functions of multiple arguments and pairing is not used is the basic reason that we feel free to return to the usual notation for function application.

We now show that TRC can 'encode' notions of propositional logic. The projection operators p_1 and p_2 will encode the truth values 'true' and 'false'. We prove the following lemma, which shows that the syntactical type of expressions T which represent propositions can be raised and lowered freely.

Lemma 1. *For any term T such that "$T = p_1 \vee T = p_2$" is a theorem and for any integer n, there is a term T' such that "$T = T'$" is a theorem and all instances of variables in T' occur inside a single instance of T in T' with type n.*

Proof. $p_1(p_1, p_2) = p_1$ and $p_2(p_1, p_2) = p_2$, so "$T(p_1, p_2) = T$" is a theorem. T occurs in $T(p_1, p_2)$ with type 1, and any subterm containing a variable is clearly in T. Iterate this construction to verify the lemma for $n > 0$. $\mathrm{Eq}(K[p_1], K[p_1]) = p_1$ and $\mathrm{Eq}(K[p_2], K[p_1]) = p_2$, so "$\mathrm{Eq}(K[T], K[p_1]) = T$" is a theorem. T occurs in $\mathrm{Eq}(K[T], K[p_1])$ with type -1, and any instance of a variable here is clearly in the instance of T. Iterate this construction to verify the lemma for $n < 0$. The lemma is trivially true for $n = 0$. The proof of Lemma 1 is complete. □

It is a consequence of Lemma 1 that Theorem 1 and its Corollary can be generalized in the case where T is provably p_1 or p_2 to the situation where x (or x and y) occur with no type other than a fixed n, not necessarily 0. Use the Lemma to raise or lower the types of all the occurrences of x to 0.

Definition. We call a term f of TRC a 'characteristic function term' if "$(\forall x)(f(x)=p_1 \vee f(x)=p_2)$" is a theorem ($x$ a variable not occurring in f). We say that a formula P is 'represented by the characteristic function f relative to the variable x' if f does not contain x and "$f(x)=p_1 \leftrightarrow P$" is a theorem. If two terms represent the same formula relative to the same variable, they are provably equal; we will use the notation $\{x \mid P\}$ to represent the function which represents P relative to x, if there is one. The coincidence with set notation in NF(U) is deliberate; when both theories are being considered, we will indicate which meaning is intended.

Lemma 2. *A formula "$f=g$" can be represented by a characteristic function h relative to x if the variable x occurs in (f, g) with no type other than a fixed n. The types of all instances of variables in h will be shifted by a constant amount from their types in (f, g).*

Proof. Apply Theorem 1 and Lemma 1 to the term Eq(f, g)—$h=(\mathsf{L}x)$(Eq(f, g)), where Eq(f, g) may need to be raised or lowered in type in order to apply Theorem 1. Note that if x does not occur in f or g. The type of Eq(f, g) can actually be manipulated freely in $(\mathsf{L}x)(\text{Eq}(f, g)) = K[\text{Eq}(f, g)]$. The proof of Lemma 2 is complete. □

Lemma 3. *If the formulae P and Q are represented by characteristic functions f and g relative to the variable x, then a formula "$\sim P$" or "P & Q" is represented by a characteristic function h relative to x, all instances of variables in f or g implying instances of the same type in h.*

Proof. For "$\sim P$", let $h=(\mathsf{L}x)(\text{Eq}(f(x), p_2))$. For "$P$ & Q", let $h=(\mathsf{L}x)(\text{Eq}((f(x), g(x)), (p_1, p_1)))$. The rest follows from Theorem 1. The proof of Lemma 3 is complete. □

Lemma 4. *If P is represented relative to x by a characteristic function f which contains the variable y with no type other than a fixed n, the formula "$(\forall y)(P)$" is represented by a characteristic function h relative to x, in which all variables other than y (which does not appear) appear with the same types with which they appear in f.*

Proof. Let $h=(\mathsf{L}x)(\text{Eq}((\mathsf{L}y)(f(x)), K[p_1]))$. The rest follows from Theorem 1 and Lemma 1. The proof of Lemma 4 is complete. □

Theorem 2. *Let P be a formula of* TRC *which is 'stratified' in the following sense: integer 'types' can be assigned to each variable in such a way that for each subformula "$f=g$", there is a constant such that each variable appears in (f, g) with no other type than its 'type' plus the constant. Then P is represented by a characteristic function relative to each variable x.*

Proof. We proceed by induction on the structure of formulae. We can choose the typing scheme so that x is given type -1. Lemma 2 gives us the result for atomic formulae — certainly x will appear with a fixed type in an atomic formula. We follow the rule when applying Lemma 2 that the type with which each variable appears in the characteristic function used to represent an atomic formula will be its 'type'. If x actually appears in the atomic formula, we are forced to do this by the construction (the type of x is first raised or lowered to 0 so that Theorem 1 can be applied, then lowered by one to -1 by the application of Theorem 1); if x does not appear, we are free to do this because we can manipulate types freely. If the type with which each variable appears in functions representing subformulae Q and R is its 'type', Lemma 3 allows us to represent subformulae "$\sim Q$" and "$Q \& R$" while preserving the condition (types of variables are fixed by the construction of Lemma 3). The same holds for Lemma 4 — the condition that the type of y is its 'type' wherever it appears enables us to apply the Lemma to represent subformulae "$(\forall y)(Q)$", and the construction fixes the types of variables, so the condition is preserved. The proof of Theorem 2 is complete. □

Corollary. NFU *can be interpreted in* TRC.

Proof. Define NFU$\vdash x \in y$ as "$y(x) = p_1 \& (\forall z)(y(z) = p_1 \vee y(z) = p_2)$". Represent concepts of first-order logic with equality in NFU by the same concepts in TRC. Observe that the 'types' in the sense of Theorem 2 of x and y in "$x = y$" are the same, and that the 'type' of y in "$x \in y$" is one higher than the type of x. Thus, the stratification conditions for the existence of a set $\{x \mid P\}$ in NFU correspond exactly to the conditions for the existence of a characteristic function $\{x \mid P\}$ representing the corresponding formula in TRC relative to x. Moreover, the formula "$x \in \{x \mid P\}$" of TRC is logically equivalent to P by Theorem 2 and the fact that $\{x \mid P\}$ is a characteristic function, for stratified P. Observe that two objects of TRC have the same 'elements' if they are characteristic functions with the same extension, thus equal, or if they are both $K[p_2]$ or not a characteristic function, in which case they both have no 'elements'. The Axioms of Comprehension and Extensionality of NFU hold in our interpretation of NFU in TRC. The proof of the Corollary is complete. □

We can do better than this. We can actually interpret NF in TRC. To do this, we need to change our interpretation of sets; the natural interpretation of a set as a function from the universe to the truth values needs to be modified. We give the construction as a theorem.

Theorem 3. *Sentences of* NF *can be translated into sentences of* TRC *in such a way that theorems translate to theorems. Thus, the consistency of* TRC *implies the consistency of* NF.

Proof. The idea is to construct a bijection Push from functions of TRC onto characteristic functions of TRC, then to define $\mathrm{NF} \vDash x \in y$ as "$\mathrm{Push}(y)(x) = p_1$". Once any such bijection Push is constructed, the Axiom of Comprehension of NF is verified, using Theorem 2 to verify that collections defined in terms of this 'membership' in a stratified manner have characteristic functions, then observing that the inverse image of such a characteristic function under Push has the intended members with this definition of membership. The Axiom of Extensionality follows because functions f and g have the same elements exactly if $\mathrm{Push}(f)$ and $\mathrm{Push}(g)$ have the same extension, so are equal, and Push is a bijection, so $f = g$ iff $\mathrm{Push}(f) = \mathrm{Push}(g)$. This is a variation on the permutation method for relative consistency and independence proofs in NF, introduced by Dana Scott in [13] (see also [7], [8]), which we use several times below. Maurice Boffa used essentially the same technique in [1] to show that NF could be interpreted in NFU + "the cardinality of the set of sets is equal to the cardinality of the universe".

It remains to construct such a bijection Push. The bijection Push which we will use sends each function which is not a characteristic function to the characteristic function of the set of ordered pairs which is usually thought of as that function, and sends this characteristic function to its image under the same operation, and so forth, fixing all functions which are not images of non-characteristic functions under iterated applications of the operation. We reiterate that any bijection from the universe onto characteristic functions can actually be used to define a model of NF.

We define some objects of TRC. Char is defined as $\{f \mid (\forall x)(f(x) = p_1 \vee f(x) = p_2)\}$. Setof is defined as $(\mathsf{L}f)(\{x \mid f(p_1(x)) = p_2(x)\})$. Note that Setof is a one-to-one map. Pushset is defined as $\{x \mid (\forall a)((\forall y)((\mathrm{Char}(y) = p_2 \rightarrow a(y) = p_1) \mathbin{\&} (a(y) = p_1 \rightarrow a(\mathrm{Setof}(y)) = p_1)) \rightarrow a(x) = p_1)\}$. Pushset represents the closure of the class of non-characteristic functions under iterated application of Setof. Push is defined as $(\mathsf{L}x)(\mathrm{Pushset}(x)(\mathrm{Setof}(x), x))$ (note that the type of $\mathrm{Pushset}(x)$ can be lowered to recover stratification, since Pushset is a characteristic function term). It is straightforward to verify that Push is a bijection from the universe of TRC to the characteristic functions of TRC. The proof of Theorem 3 is complete. □

Part 3. Theories with ur-elements

In this part, we first define the theory TRCU, then show how to interpret NFU + Infinity + "there is a surjective pairing function on the universe" in any model of NFU + Infinity.

Since NFU is known to be consistent, a natural way to try to construct a theory similar to TRC for which we can find a relative consistency proof is to define a theory TRCU with 'ur-elements'—with some kind of weakening of the axiom of extensionality. It is not immediately clear what an ur-element is in the context of a theory of functions and the identification of pairs and products of functions stipulated in Axiom IV complicates the definition of TRCU somewhat.

Terms of TRCU are the same as terms of TRC, except that there is an additional atomic term ?, which may be thought of as a marker for undefined values. The additional axiom for the atom ? asserts that $K[?] = (?, ?) = ?$. Axiom VII is replaced by two axioms, whose statement requires that we define the concept of 'iterated projection': the set of iterated projections of an object x is the minimal set containing x and closed under application of p_1 and p_2. We call an object u an 'ur-element' if $u(x) = ?$ for all x and $u \neq ?$. The first axiom of extensionality of TRCU asserts that if $f(x) = g(x)$ for all x and no iterated projection of f or g is an ur-element, then $f = g$. The second axiom of extensionality asserts that no iterated projection of an object $K[f]$, Abst(f), or Abst(f)(g) is an ur-element. The second axiom of extensionality ensures that the two special cases of extensionality needed in the proof of Theorem 1 hold in TRCU. It also ensures that for any term T, $(\mathsf{L}x)(T)$ as defined in the proof of Theorem 1 has no iterated term which is an ur-element; this implies that if "$T = U$" is a theorem, "$(\mathsf{L}x)(T) = (\mathsf{L}x)(U)$" is a theorem (this is an easy consequence of extensionality in TRC), which has important consequences for the character of TRCU as a system of combinatory logic. Theorem 2 holds for TRCU, and NFU can be interpreted in TRCU as described after the proof of Theorem 2. The proof of Theorem 3 fails in TRCU. TRCU can be interpreted in NF, taking a little care to ensure that an object with the properties of ? is present. TRCU cannot be interpreted in NFU without additional assumptions, since the presence of surjective pairing in TRCU implies that the universe is infinite, which cannot be proven in NFU. We will show that TRCU can be interpreted in NFU with the Axiom of Infinity.

The motivation for the definition of TRCU is as follows: ur-elements should be thought of as objects which are not functions — when one applies them to something, one gets the 'error' value ? (which is to be considered a function, just as the empty set can still be considered a set in NFU). The simple assertion of the existence of ur-elements is not sufficient as a weakening of extensionality, because of Axiom IV: if u is an ur-element, and f is a function, (u, f) and $(?, f)$ have the same extension, but neither is an ur-element. Preservation of Axiom IV motivates the use of the concept of iterated projection in the statement of the first axiom of extensionality. Note that the concept of iterated projection can be encoded in the language of TRC by the use of the techniques of Theorem 2; the axiom can in principle be stated without recourse to English. The second axiom of extensionality is motivated by the need to preserve the provability of Theorem 1; it also preserves the character of TRCU as a 'lambda-calculus' by ensuring that substitutions of equals for equals in T in expressions $(\mathsf{L}x)(T)$ are permissible. (The theory TRCL in which we adjoin to the axioms of TRC without extensionality the existence of an atom L such that if $f(x) = g(x)$ for all x,

$L(f) = L(g)$, and $L(L(f)) = L(f)$ [L is a retraction to canonical functions of each extension] is actually also adequate to interpret NFU + Infinity; it is convenient to add axioms with the effect that L fixes terms $(\mathrm{L}x)(T)$: $L(\mathrm{Id}) = L$, $L(K[f]) = K[f]$, $L(f, g) = (L(f), L(g))$, and $L(\mathrm{Abst}(f)(g)) = \mathrm{Abst}(f)(g)$; in this case, L is definable as $(\mathrm{L}f)((\mathrm{L}x)(f(x)))$ and $L(L(f)) = L(f)$ becomes a theorem).

We summarize the additional axioms of TRCU:

O. $K[?] = (?, ?) = ?$.

VIIA. If $f(x) = g(x)$ for all x, then $f = g$ or some iterated projection of f or g is an ur-element and distinct from the corresponding iterated projection of g or f.

VIIB. No iterated projection of $K[f]$, $\mathrm{Abst}(f)$, or $\mathrm{Abst}(f)(g)$ is an ur-element, for any f or g.

We now construct a model of NFU with desirable properties within the theory NFU with the Axiom of Infinity, which was shown to be consistent by Jensen in [11]. We distinguish between 'sets' and 'ur-elements' in NFU as follows: all objects with elements and one distinguished object { }, the empty set, without elements, are taken to be 'sets', and all other objects are taken to be 'ur-elements'. The elements of the model NFU2 which we will construct are the 'sets of sets' — those 'sets' all of whose elements are 'sets'. We define NFU2 $\models a \in b$ as '$a \in b$ and all elements of b are sets of sets" in the sense of the underlying NFU + Infinity. It should be clear the Axiom of Extensionality of NFU is satisfied in the model NFU2, with those sets of sets which have elements which are not sets of sets being interpreted as ur-elements in the model. Note that the type of a is one less than the type of b in the formula defining membership in the model; thus, the stratification conditions for the existence of a collection of elements of NFU2 defined in terms of first-order logic, equality and 'membership' are the same as the stratification conditions for the existence of the set in the underlying NFU defined in the analogous way using the original membership and universe of objects. Since a collection of elements of NFU2 is itself an element of NFU2 and has the same elements in the sense of NFU2 as it does in the sense of the underlying NFU, this implies that the Axiom of Comprehension of NFU holds in NFU2. We have established that we have constructed a model of NFU — it is straightforward to establish that it is a model of NFU + Infinity.

In NFU + Infinity, the Quine ordered pair can be defined and is a surjective pairing relation on sets of sets — thus, the model NFU2 we have constructed has a surjective pairing relation on its universe analogous to the Quine ordered pair. The Quine ordered pair is a manipulation of 'elements of elements' of objects — other manipulations of elements of elements of objects in NF can be emulated in this model of NFU + Infinity by manipulating elements of elements of objects in the model, not in the sense of the model, but in the sense of the underlying model of NFU + Infinity.

Part 4. Interpreting TRC(U) in NF(U)

It remains to show that the converse of Theorem 3 is also true; it is possible to interpret TRC in NF. We will then show, in an analogous but somewhat more complicated manner, that TRCU can be interpreted in the model of NFU + Infinity whose elements are the 'sets of sets' in a general model of NFU + Infinity, and so is consistent relative to the usual set theory by Jensen's results of [11]. In [9], we showed how to construct models of TRCU directly in the usual set theory.

To do this, we need to define function application in NF(U) in such a way that every set is interpreted as a function of universal domain. The natural definition of the parameterized proper name "$f(x)$" in NF(U) is "the object y such that f is a function and $(x, y) \in f$", where the natural definition of "f is a function" is "for each x, there is no more than one y such that $(x, y) \in f$", and "$f(x)$" is taken to be the empty set when its description does not make sense, as usual. Our first approximation to function application will be a slight modification of this: initially, we shall define "f is a function" as "for each x, there is no more than one y such that $(x, y) \in f$, and there is no x such that $(x, \{\})\} \in f$", excluding $\{\}$ from the range of results of function application except as a default value, and define "$f(x)$" as above with the new definition of 'function' replacing the old.

The second approximation to function application is related to the first much as the definition of set membership interpreting NF in TRC is related to the definition of set membership interpreting NFU. We take the set Char to be the set of sets all of whose elements are of the form $(a, \{\{\}\})$; each element of Char is currently interpreted as a function taking elements of a certain set to $\{\{\}\}$ and elements of the complement of that set to $\{\}$ — a 'characteristic function' of some set. The function Charof takes each set A to the element of Char consisting of those pairs $(a, \{\{\}\})$ such that $a \in A$ — the 'characteristic function' of A. We define Pushset as the inductive closure of the complement of the set of functions under the function Charof, then define Push as the map which takes each element of Pushset to its image under Charof and each element of the complement of Pushset to itself. It is straightforward to verify that these sets and functions have stratified definitions and that Push is a bijection from sets onto functions (in NF, a bijection from the entire universe onto functions).

Our second approximation to the definition of function application is as follows: we define "$f(x)$" as "the object y such that f is a set and $(x, y) \in$ Push(f)", using the previous definition of function application to interpret "Push(f)". In NF, where each object is a set, this gives an extensional theory of functions; in NF(U) each ur-element, as well as $\{\}$, has the value $\{\}$ everywhere (recalling that an object defined by description is taken to be $\{\}$ when the description fails). Note that we are again using a variation on the permutation method of [13].

We show that NF(U) as a theory of functions has an abstraction property: if T is a parameterized proper name in which x appears with no type other than the type of T, the set of pairs (x, T) such that $T \neq \{\ \}$ has a stratified definition, and its inverse image under Push, which we will call $(\mathsf{L}_1 x)(T)$, has the property that "$(\mathsf{L}_1 x)(T)(x) = T$" is a theorem. (Note that Rosser defined 'lambda-abstraction' in NF in [12, p. 325]). We could use this 'lambda-abstraction' over NF to define p_1, p_2, Abst, and Eq in such a way as to obtain a theory satisfying all the axioms of TRC except Axiom IV: we define "p_i" as $(\mathsf{L}_1 x)(p_i(x))$ [where the expression abstracted from is the projection as defined originally, without reference to application], "Eq" as

$$(\mathsf{L}_1 x)((Ty)(p_1(x) = p_2(x) \,\&\, y = \text{“}p_1\text{”} \vee p_1(x) \neq p_2(x) \,\&\, y = \text{“}p_2\text{”})),$$

$K[f]$ as $(\mathsf{L}_1 x)(f)$, and "Abst" as $(\mathsf{L}_1 f)((\mathsf{L}_1 g)((\mathsf{L}_1 h)(f(K[h])(g(h)))))$. It is not the case that $(f, g)(h) = (f(h), g(h))$ can be expected to hold in NF with our definitions of pairing and application. We can define an additional 'atom' Prod so as to satisfy an Axiom IV′: $\text{Prod}(f, g)(h) = (f(h), g(h))$:

$$\text{Prod} = (\mathsf{L}_1 x)((\mathsf{L}_1 y)(p_1(x)(y), p_2(x)(y))).$$

The resulting theory satisfies results analogous to those proven for TRC above; it has an abstraction theorem, encodes propositional logic, and interprets NF. The theory obtained in the same way from NFU has the property that if f and g have the same extension but are not equal, one must be an ur-element and the other must be an ur-element or $\{\ \}$, and that any pair one of whose terms is an ur-element is an ur-element. This is quite different from the treatment of ur-elements in TRCU as defined above.

The complications in our constructions arise from getting Axiom IV to hold, as well as the first axiom of extensionality of TRCU. This will require a careful analysis of the Quine ordered pair. Recall that in the case of NFU + Infinity, the Quine ordered pair is defined on 'sets of sets' in a larger model of NFU + Infinity, and that these 'sets of sets' make up the whole of the model of NFU + Infinity in which we will interpret TRCU.

The Quine ordered pair (x, y) is defined as

$$\{z \mid (\exists w)((S[w] = z \,\&\, w \in x) \vee (S[w] + \{0\} = z \,\&\, w \in y))\}.$$

The projection $p_1(x)$ is defined as $\{z \mid S[z] \in x\}$ and $p_2(x)$ is defined as $\{z \mid S[z] + \{0\} \in x\}$. It is easy to see that functions p_1 and p_2 exist under our definition of function application. Iterated projection operators (compositions of p_1 and p_2) are of the form $P(x) = \{z \mid nS[z] + A \in x\}$, where n is the number of projections composed, $nS[z]$ represents the image of z under the function which adds n to each number and fixes each non-number, and A is a set of numbers less than n which encodes the series of projections composed. We say that distinct projection operators P and Q are 'independent' if there is no projection operator R such that R composed with P is Q or R composed with Q is P. It is obviously possible to realize any assignment of values to a finite collection of pairwise independent iterated projections of an object, under any definition of the ordered pair; a special property of the Quine ordered pair is that it is possible to realize any assignment of values to any collection of pairwise independent iterated projections of an object. Two projection operators are independent iff neither of their associated subsets of $\mathbb{N}$ is a subset of the other. Given an assignment of values to a collection of pairwise independent projection operators of an object, we take each intended projection, shift numbers which are elements of elements as indicated by the number associated with the projection, and add as elements of elements the elements of the finite set associated with the projection. The condition of pairwise independence ensures that the collection of modified intended projections will be disjoint; its union is an object with the intended projections. Note that any collection of assignments of values to iterated projections any finite subset of which can be realized can be realized in this way; when projections in such a set are not independent, so that we have some P and Q with P equal to the composition of R and Q, the assignment to P is completely determined by the assignment to Q and can be omitted; take the projection operators with minimal associated sets and numbers — these will make up a pairwise independent collection of projections, and assignments to these completely determine the original assignments.

We wish to define a bijection between sets of sets and sets of sets which have no elements of elements which are natural numbers. We observe that no number is a pair of natural numbers, and define Npush as the map which takes each number n to $(n, 0)$, each pair of numbers (m, n) to $(m, n+1)$, and every other object to itself. We define Npurge as the map which takes a set of sets to the set of sets obtained by replacing each element of an element with its image under Npush: Npurge is a bijection from sets of sets onto sets of sets having no natural numbers as elements of elements. We define Nrestore as the inverse of Npurge.

We define a type-raising operation on sets of natural numbers. If n is a natural number, let Tn be the cardinality of the singleton image of an element of n; if A is a set of natural numbers, let $T[A]$ be $\{Tn \mid n \in A\}$. Let $U[A]$ be the unique B such that $T[B] = A$; let $2U[A]$ be $U[U[A]]$. All of these operations would be trivial in the presence of the Axiom of Counting. Observe that these type raising operations induce a type raising operation on iterated projection operators: if P is the projection determined by the finite set A and the number n, let $T[P]$ be the projection determined by $T[A]$ and Tn.

We can now define 'infinite iterated projection operators' by analogy with the finite compositions of projection operators, as follows: when A is a set of natural numbers, define pA as the function which takes x to Nrestore($\{z \mid \mathbb{N}z = \{\,\} \;\&\; z + A \in x\}$). If we did not apply Nrestore, $pA(x)$ would have no elements of elements which were natural numbers. Note that the relative type of pA is two higher than the relative type of A. Note also that

$$pA(p_1(x)) = p[S[A]](x) \quad \text{and} \quad pA(p_2(x)) = p[S[A] + \{0\}](x).$$

We define Product(f) as the unique object such that

$$pA(\text{Product}(f)(x)) = f(T[A]).$$

The motivation of the definition is to construct a product operation corresponding to the projections pA; the complexities of the definition (the fact that Product(f) is the constant function of the object actually desired and the appearance of the T operation) arise from stratification requirements. We can now prove a theorem.

Theorem 4. TRC *can be interpreted in* NF; *thus,* TRC *and* NF *have exactly the same consistency strength and expressive power.*

Proof. Redefine "$f(x)$" as Product$((\mathsf{L}_1 A)(pA(f)(x)))(y)$, using the previous definition of function application to interpret the expression on the right. Note that the image of "$f(x)$" under pA will be $p[T[A]](f)(x)$ in the old sense. It is straightforward to check that stratification requirements are satisfied.

Define a lambda-abstraction operation for the new definition of function application as follows: $pA((\mathsf{L}x)(T)(x))$ needs to be $pA(T)$, so $p[T[A]]((\mathsf{L}x)(T))$ needs to be $(\mathsf{L}_1 x)(pA(T))$, so $pA((\mathsf{L}x)(T))$ is $(\mathsf{L}_1 x)(p[U[A]](T))$, and $(\mathsf{L}x)(T)$ is Product$((\mathsf{L}_1 A)((\mathsf{L}_1 x)(p[2U[A]](T))))(y)$. It is straightforward to verify that such an object exists and that " "$(\mathsf{L}x)(T)(x)$" = T" is a theorem (quotes indicating use of new function application) if T contains x with no type other than the type of T.

We then use the new definition of 'lambda-abstraction' to define objects "p_1", "p_2", "$K[f]$" [for each f], "Abst", and "Eq", by abstraction over appropriate terms (using the original pair as the pair of TRC). It is straightforward to verify each axiom of TRC other than Axiom IV for the resulting theory.

We consider "$(f, g)(h)$" in the new sense. $pA(\text{"}(f, g)(h)\text{"}) = p[T[A]](f, g)(h)$, so

$$p[S[A]](\text{"}(f, g)(h)\text{"}) = p[[T[S[A]]](f, g)(h) = p[[S[T[A]]](f, g)(h)$$
$$= p[T[A]](f)(h) = pA(\text{"}f(h)\text{"}) = p[S[A]](\text{"}f(h)\text{"}, \text{"}g(h)\text{"}).$$

Similarly

$$p[S[A] + \{0\}](\text{"}(f, g)(h)\text{"}) = pA(\text{"}g(h)\text{"})$$
$$= p[S[A] + \{0\}](\text{"}f(h)\text{"}, \text{"}g(h)\text{"}).$$

Thus, the interpretations of expressions on each side of Axiom IV have the same images under each infinite iterated projection operator, and must be the same object. (It is straightforward to establish that if the construction here is carried out in NFU + Infinity, one obtains a model of the theory TRCL discussed parenthetically in Part 3). The proof of Theorem 4 is complete. □

Theorem 5. TRCU *can be interpreted in* NFU + Infinity; *thus, these theories have exactly the same consistency strength and expressive power, and* TRCU *is consistent relative to the usual set theory.*

Proof. We define an equivalence relation: f and g are equivalent if for each finite projection P there is a finite projection Q such that $Q(P(f))$ and $Q(P(g))$ are equal. We use this equivalence relation, which we call 'termwise equivalence almost everywhere', to interpret equality in a first approximation to TRCU.

We redefine function application in such a way that all functions respect termwise equivalence almost everywhere. Let Termwise(f) be the map which applies f to each infinite projection of x: $pA(\text{Termwise}(f)(x)) = f(pA(x))$. Termwise($f$) will respect equivalence for each f. Now define Pushset as the inductive closure under Termwise of the set of functions which do not respect the equivalence relation. Define Push(f) as Termwise(f) on Pushset and f elsewhere. Push is a bijection from functions onto functions which respect the equivalence relation. Redefine function application "$f(x)$" as Push(f)(x). (The use of Push ensures that the construction given here builds TRC at the next step if given NF—if we did not care about this, we could treat functions not respecting the equivalence relation as ur-elements.)

Now carry out the construction in the proof of Theorem 4, defining "$f(x)$" so that pA("$f(x)$") will be $p[T[A]](f)(x)$ in the previous sense; this definition of application respects the equivalence relation which interprets equality, as do pairing and projections, so we can define lambda-abstraction on parametrized proper names which are determined up to equivalence by the equivalence classes of included variables and verify the axioms of TRC (other than extensionality) as in the proof of Theorem 4. Failures of extensionality in this theory are associated with the presence of infinite iterated projections which are ur-elements; the operation which changes each ur-element appearing as an infinite iterated projection of its argument to $\{\,\}$ is a retraction from objects onto canonical objects with the same extension, and exists as a function because it respects stratification conditions and the equivalence relation. The new interpretation has the property that termwise equivalence almost everywhere implies equality, so an object is completely determined if some finite projection of each of its finite projections is given.

We now introduce the interpretation of TRCU, which is analogous to the interpretation of NFU in NF without any axiom of extensionality in Marcel Crabbé's [4]. The equivalence relation which will interpret equality is that of having the same extension in the previous interpretation. We refer to a function as 'useful' if it respects this equivalence relation and has all values fixed under the retraction (an object is fixed under the retraction precisely if it has no infinite iterated projection which is an ur-element). Note that if two useful functions have equivalent values everywhere, they have equal values everywhere and so are equivalent. We define "$f(x)$" as follows: for each finite projection P such that $T[P](f)$ is useful, we assign P("$f(x)$") the value $T[P](f)(x)$; for each finite projection P such that $T[P][f]$ has no finite projection which is useful, we assign P("$f(x)$") the value $\{\,\}$. Since any finite collection of such assignments can be realized, the whole collection can be realized, by properties of the Quine pair;

since we make an assignment to some finite projection of each finite projection, the assignments determine "$f(x)$" exactly. Lambda-abstraction for parameterized proper names T determined up to equivalence by the equivalence classes of included variables is defined by applying the previous definition of lambda-abstraction to the image of T under the retraction; note that all projections of a lambda-abstract will be useful. Since application, pairing, projection, and constant terms are of this sort, we can verify all axioms of TRCU except Axiom IV and the axioms of ? and extensionality. Axiom IV holds because projections of "$f(x)$" are defined in terms of corresponding (type-raised) projections of f.

? is interpreted as $\{\ \}$; satisfaction of "$(?, ?) = K[?] = ?$" is immediate. Suppose that f and g are not equivalent but have the same extension under the new definition of function application; this means that they had different extensions under the old definition. Any corresponding iterated projections of f and g which are useful have values in the old sense at every x which are equivalent and thus equal, since fixed under the retraction, so such projections of f and g are equivalent. Thus, there must be a projection P such that either $P(f)$ or $P(g)$ has no finite projection which is useful, and $P(f)$ and $P(g)$ are not equivalent (otherwise, the useful projections would completely determine the values of f and g). Such projections have value $\{\ \}$ in the new sense everywhere, but are not equivalent to $\{\ \}$; the corresponding $P(g)$ or $P(f)$ must be a distinct object of the same sort or an object equivalent to $\{\ \}$, so the first axiom of extensionality of TRCU is satisfied. The second axiom of extensionality is satisfied because any object with which it is concerned has all finite projections useful, while any object interpreted as an ur-element has a finite projection none of whose finite projections are useful. The proof of Theorem 5 is complete. □

The interpretations of TRC and TRCU developed in the proofs above have additional properties which might be desirable as axioms. In either interpreted theory, any characteristic function f which has the property that for each x there is exactly one y such that $f(x, y) = p_1$ has an associated function F such that $F(x) = y$ iff $f(x, y) = p_1$; this is certainly desirable, and does not follow from the axioms of either theory. An interpretaton of TRC(U) using the construction of Theorem 5 (which produces an interpretation of TRC if carried out in NF) satisfies the assertion that objects which are termwise equivalent almost everywhere are equal; it also has the property that any assignment of values to a (possibly infinite) collection of finite iterated projections of an object any finite subcollection of which can be realized can itself be realized, because the Quine pair has this property. In [9], we called a pairing relation with these properties a 'compact' pairing relation, and we used such pairing relations in the construction of models of TRCU in the usual set theory. We believe that compactness of the pair might be a useful axiom to adjoin to TRC(U) as well.

Part 5. Concluding remarks

We believe that NFU (and TRCU) deserve some attention as alternate foundations for mathematics. We hope that the equivalence of TRC and NF will cast some light on the problem of the consistency of NF, but we believe that the systems TRC and NF, while probably consistent and interesting in themselves, are too strange to be plausible foundations for mathematics. The insistence on strict extensionality in these theories has such odd effects as making it possible to disprove the Axiom of Choice (see Specker's [14]). The use of the Quine ordered pair was the only feature of Rosser's development of mathematical logic based on NF in [12] which could not be emulated in NFU with Infinity; he described an adequate mathematical logic based on NF with the Axiom of Counting (described above) and the Axiom of Denumerable Choice. His development can be duplicated in NFU with Counting (which implies Infinity) and the Axiom of Choice. This theory has the advantages of provable relative consistency and the presence of the unrestricted Axiom of Choice. One can construct an internal model of NFU with a surjective pairing analogous to the Quine ordered pair as we did above; it is also possible to prove the existence of a surjective pairing relation on the universe in NFU + Infinity + Choice. In the presence of an inaccessible cardinal, it is possible to present a model of NFU which contains an inner model of ZFC. Axioms can be given which are satisfied in such a model and which combine the facilities of ZFC for reasoning about 'small' collections and of NFU for reasoning about 'large' collections (see [9, pp. 63–64]). This gives a much richer theory of 'large' collections than one has in the usual extensions of ZFC with 'proper classes' at a reasonable cost in additional complexity.

We believe that TRC(U), as a combinatory logic which is compatible with logical notions yet untyped, may find some application in theoretical computer science. We have established in [9] that, with a suitable notion of reduction, TRC satisfies the Church–Rosser property and strong normalization (mod the non-constructive character of equality). TRCU can interpret typed combinatory logic easily, and can also interpret such systems as the polymorphic type theory which underlies the computer language ML in a straightforward fashion. We believe that the notion of 'set' is a less natural foundation for computer science than the notion of (potentially universally applicable) 'rule' or 'function'; type restrictions on application of functions are useful, but the fact that polymorphic systems are studied indicates that there is also a need to be able to avoid such restrictions. We can give a computer-science oriented motivation for the restrictions on abstraction in TRCU using the distinction between the performance of a stored program and its address (see [10], where we actually work mostly with TRCL).

References

[1] M. Boffa, The consistency problem for NF, J. Symbolic Logic 42 (1977) 215–220.
[2] M. Boffa, The point on Quine's NF (with a bibliography), Teoria 4 (2) (1984) 3–13.

[3] M. Boffa, ZFJ and the consistency problem for NF, in: The Kurt Gödel Gesellschaft Year Book (1988) 102–106.
[4] M. Crabbé, On NFU, Preprint.
[5] H.B. Curry and R. Feys, Combinatory Logic, Vol. I (North-Holland, Amsterdam, 1958).
[6] J. Girard, Proofs and Types (Cambridge University Press, Cambridge, 1989).
[7] C.W. Henson, Type-raising operations on cardinal and ordinal numbers in Quine's "New Foundations", J. Symbolic Logic 38 (1978) 59–68.
[8] C.W. Henson, Permutation methods applied to Quine's "New Foundations", J. Symbolic Logic 38 (1973) 69–76.
[9] M.R. Holmes, Systems of Combinatory Logic related to Quine's "New Foundations", Ph.D. Thesis, SUNY Binghamton, 1990.
[10] M.R. Holmes, Systems of combinatory logic related to Quine's "New Foundations" applied to program semantics, Preprint.
[11] R.B. Jensen, On the consistency of a slight (?) modification of Quine's "New Foundations", Synthése 19 (1969) 250–263.
[12] J.B. Rosser, Logic for Mathematicians, second edition (Chelsea, New York, 1978).
[13] D. Scott, Quine's individuals, in: Logic, Methodology and Philosophy of Science, Proc. of the 1960 Intl. Congress, (Stanford University Press, Stanford, CA, 1962) 111–115.
[14] E.P. Specker, The axiom of choice in Quine's "New Foundations for Mathematical Logic", Proc. Nat. Acad. Sci. USA 39 (1953) 972–975.
[15] E.P. Specker, 1962. Typical ambiguity, in: Logic, Methodology and Philosophy of Science, Proc. of the 1960 Intl. Congress (Stanford University Press, Stanford, CA, 1962) 116–124.
[16] W.V. Quine, New Foundations for Mathematical Logic, Amer. Math. Monthly 44 (1937) 70–80.
[17] W.V. Quine, On ordered pairs, J. Symbolic Logic 10 (1945) 95–96.
[18] W.V. Quine, Mathematical Logic, revised edition, (Harvard University Press, Cambridge, MA, 1951).