Systems of Combinatory Logic Related to Predicative and "Mildly Impredicative" Fragments of Quine's "New Foundations"

M. Randall Holmes

April 22, 2020

Contents

1 Introduction

In [1], we introduced a system *TRC* of illative combinatory logic precisely equivalent in consistency strength and expressive power to Quine's set theory "New Foundations" of [2] (usually called *NF*). It remains an open question whether *NF* is consistent (relative to the usual set theory). We also exhibited a system of combinatory logic *TRCU*, a weakening of *TRC*, which is precisely equivalent in consistency strength and expressive power to Jensen's *NFU* ("New Foundations with 'ur-elements"') with the addition of the Axiom of Infinity. Jensen showed in [3] that this theory is consistent relative to the usual set theory.

Marcel Crabbé introduced fragments *NFP* (predicative *NF*) and *NFI* (which we call "mildly impredicative" *NF*) of *NF* in [4], and showed them to be consistent and of quite low consistency strength. We will present a fragment *TRCI* of *TRC* which is precisely equivalent in consistency strength and expressive power to *NFI*; we also present a weaker fragment *TRCP* of *TRC* which is related to

NFP but for which we have not been able to establish an equivalence with *NFP*. The fragment *TRCP* related to *NFP* commands interest because it is "natural" in character, whether an exact equivalence can be established or not.

We review the definition of *NF*, and give the definitions of the fragments *NFU*, *NFP*, and *NFI*. *NF* is the first-order theory with equality and membership whose non-logical axioms are extensionality (sets with the same elements are equal), and those instances of the axiom scheme of comprehension (for each condition ϕ in the language of *NF* and variable x, "$\{x|\phi\}$ exists") in which the condition ϕ is "stratified". A formula in the language of *NF* is said to be "stratified" if each variable occurring in the formula can be assigned a non-negative integer type in such a way as to obtain a formula of the simple theory of types.

NFU is obtained from *NF* by weakening extensionality so that if two objects have the same elements, either they are equal or both have no elements. This amounts to the introduction of atoms or ur-elements. *NFP* and *NFI* have full extensionality, but weaker schemes of comprehension. The formula meaning "$\{x|\phi\}$ exists" belongs to the comprehension scheme of *NFI* if it is stratified and satisfies the further restriction that the assignment of types can be made in such way that types more than one greater than the type of x (i. e., greater than the intended type of $\{x|\phi\}$ itself) do not occur. We refer to the condition ϕ as being "mildly impredicative" (relative to x). It belongs to the comprehension scheme of *NFP* if it satisfies the further restriction that no bound variable in ϕ can have type greater than the type of x; all variables of the exact type of $\{x|\phi\}$ are parameters. We refer to such a condition ϕ as "predicative" (relative to x).

Con(*NFU*) is a theorem of Peano arithmetic; *NFU* + Infinity has the exact consistency strength of the simple theory of types with the Axiom of Infinity (see [3]). Crabbé showed in [4] that *NFI* interprets second-order arithmetic, while Con(*NFI*) is a theorem of third-order arithmetic, and that *NFP* interprets Robinson's arithmetic, while Con(*NFP*) is a theorem of Peano's arithmetic. In [5], we have shown how to construct a model of *NFI* within which it is possible to interpret nth order arithmetic for every n. *NF* is known to be at least as strong as the simple theory of types with the Axiom of Infinity; the proof of the Axiom of Infinity in *NF* was first achieved by Specker, who showed in [6] that the Axiom of Choice is false in *NF*–since the universe is a set in *NF*, and cannot be well-ordered (by $\sim AC$), it cannot be finite. No one has succeeded thus far in constructing a model of *NF* in a more familiar set theory or in deriving a contradiction from *NF*. The Axiom of Choice is consistent with *NFP*, *NFI*, and *NFU* + Infinity.

These set theories have the strange feature that "very large" collections, such as the universe, are sets. The natural numbers can be defined, using Frege's original definition, in such a way that the natural number n is the set of all sets with n elements (for concrete n). The set of natural numbers N can then be defined (in *NF*, *NFI*, *NFU*) as the set of all objects which belong to each set which contains 0 and is closed under the "successor" operation. In *NFP*, this definition fails, as it involves a reference to all inductive sets, and the inductive sets would be assigned the same type as the set being defined. For a fuller treatment of set-theoretical constructions in *NF*, see [1]; Rosser's [7] is a full-scale development of mathematical logic in an extension of *NF* (which can be adapted to *NFU* + Infinity, so avoiding the consistency problem).

NFP or *NFI* is strengthened to full *NF* by the addition of the Axiom of Union. Consider the set of all n-fold iterated singletons of objects satisfying a condition ϕ; if n is taken to be large enough, the type of the set being constructed will exceed any type used in ϕ. Now n iterated unions will give the set of objects satisfying ϕ. *NFP* is strengthened to *NFI* by the addition of the axiom scheme consisting of all assertions "The union of $\{x|\phi\}$ exists", where "$\{x|\phi\}$ exists" is an axiom of *NFP* containing no parameters of the same type as $\{x|\phi\}$. Certainly, if "$\{x|\phi\}$ exists" is an axiom of *NFI*, the assertion "$\{\{x\}|\phi\}$ exists" can be expressed by an axiom of *NFP* with no parameters of the highest type, and the union of $\{\{x\}|\phi\}$ would be $\{x|\phi\}$ if it existed, so the extension of *NFP* implies *NFI*. It is easy to establish that all the axioms added to *NFP* are actually axioms of *NFI*. Specker's proof of the Axiom of Infinity can be used to prove the Axiom of Infinity in *NFP* (and thus in *NFI*); if the Axiom of Union is false, Infinity certainly holds (finite sets have unions); if the Axiom of Union is true, we are in *NF*, and Specker's proof goes through.

2 *TRCP* Introduced

We describe the system of combinatory logic *TRCP* which we will show to be related to *NFP*. *TRCP* is a first-order theory with equality. Atomic terms of *TRCP* are Eq, Comp, π_1, π_2, and variables. If f and g are terms of *TRCP*, $f(g)$, (f,g), and $\mathrm{K}[f]$ are terms of *TRCP*. These term constructions signify function application, pairing, and the construction of constant functions, respectively. We will always write $f(g,h)$ instead of $f((g,h))$. We define Id as (π_1,π_2). We will write $\mathrm{K}^n[f]$ to represent the result of n iterated applications of the K-constructor to f. The non-logical axioms of *TRCP* are as follows:

(Const): $\mathrm{K}[f](g) = f$

(Proj): $\pi_i(f_1, f_2) = f_i$ (for i = 1,2)

(Surj): $(\pi_1(f), \pi_2(f)) = f$

(Prod): $(f, g)(h) = (f(h), g(h))$

(Comp): $\mathrm{Comp}(f, g)(h) = f(g(h))$

(Eq): $\mathrm{Eq}(f, g) =$ if $f = g$ then π_1 else π_2

(Ext): if $f(x) = g(x)$ for all x, then $f = g$

(Nontriv): $\pi_1 \neq \pi_2$

The theory *TRC* shown in [1] to be equivalent to *NF* has an atom Abst in place of Comp (Comp can be defined in *TRC*) with the more complicated axiom (Abst): $\text{Abst}(f)(g)(h) = f(\text{K}[h])(g(h))$. Note that the proposition (Id): $\text{Id}(x) = x$ follows from (Proj) and (Surj). The theory *TRCI* is described below.

We define a notion of "relative type" for subterms of a term of *TRCP*. The type of a term relative to itself is 0. If the relative type of a subterm (f, g) is n, the relative types of the obvious instances of f and g are also n. If the relative type of a subterm $f(g)$ is n, the relative type of the obvious instance of f is $n+1$ and the relative type of the obvious instance of g is n. If the relative type of an instance of $\text{K}[f]$ is n, the relative type of the obvious instance of f is $n-1$. We use this notion of relative type in the statement of an

Abstraction Theorem for *TRCP*: Let T be a term in the language of *TRCP* and let x be a variable which does not occur in T as a subterm of any subterm $\text{K}[S]$ or with any relative type other than 0. It follows that there is a term $(\lambda x)(T)$ in which the variable x does not occur such that "$(\lambda x)(T)(x) = T$" is a theorem of *TRCP*. For any variable y which occurs with type n in T and is not x, y occurs with type $n-1$ in $(\lambda x)(T)$.

Proof of the Abstraction Theorem for *TRCP*: Use induction on the structure of T. $(\lambda x)(x) = \text{Id}$. $(\lambda x)(A) = \text{K}[A]$, where A does not contain x. $(\lambda x)(U, V) = ((\lambda x)(U),(\lambda x)(V))$ (by (Prod) and ind. hyp.). $(\lambda x)(U(V)) = \text{Comp}(U, (\lambda x)(V))$, since x cannot occur in U (it would have to have type -1 in U, and there is no way for this to happen without x occurring in a subterm $\text{K}[f]$ of U). $(\lambda x)(\text{K}[U]) = \text{K}[\text{K}[U]]$; again, U cannot contain x. The assertion about types is straightforward to verify. The proof of the Abstraction Theorem is complete. ∎

Corollary: Let T be a term of the language of *TRCP* and let x and y be variables satisfying the conditions satisfied by x in the Theorem. It follows that there is a term $(\lambda xy)(T)$ not containing x or y such that "$(\lambda xy)(T)(x, y) = T$" is a theorem of *TRCP*.

Proof of the Corollary: Just as above, except that $(\lambda xy)(x) = \pi_1$; $(\lambda xy)(y) = \pi_2$; all references to "not containing x" are replaced with references to "not containing x or y". ∎

Using the Abstraction Theorem and Corollary, it is easy to show that *TRCP* is equivalent to a λ-calculus which we now define. The atomic terms of this λ-calculus are Eq and variables. If f and g are terms of the λ-calculus, $f(g)$ and (f, g) are terms of the λ-calculus. We define relative type of subterms of a term as in *TRCP* for these term constructions, and declare the relative type of f to be $n-1$ if the relative type of $(\lambda xy)(f)$ is n. If f is a term of this λ-calculus and x,y are variables which do not occur in f as subterms of terms $(\lambda zw)(g)$ or with relative type other than 0, then $(\lambda xy)(f)$ is a term of this λ-calculus.

We define $(\lambda x)(T)$ as $(\lambda yz)(T')$, where T' is the result of replacing x with (y,z) in T, y and z not occurring in T. The advantage of the use of $(\lambda xy)(f)$ as the primitive form is that we do not need primitive notions or axioms of projection. Axioms (Const) (Proj), and (Abst) of *TRCP* are replaced with the axiom scheme "$(\lambda xy)(T)(x,y) = T$"; axioms which contain π_1 and π_2 replace these with $(\lambda xy)(x)$ and $(\lambda xy)(y)$ respectively. It is clear that axioms (Const), (Proj), and (Abst) are special cases of the axiom scheme provided, with suitable definitions of π_1, π_2, Comp, and K$[f]$; the Corollary to the Abstraction Theorem shows that interpretations of all instances of this axiom scheme follow from the axioms of *TRCP*.

We now note that it is "almost" possible to interpret *TRCP* in *NFP*. The (failed) argument for this is essentially the same as the (complex) argument for the interpretation of *TRC* in *NF* given in [1]. The constructions in this argument are clearly valid in *NFI*; one apparent problem with validity in *NFP* is the occurrence of references to the set of natural numbers in the construction of the ordered pair of Quine and similar constructions. However, the relative types of natural numbers used in the construction of the Quine ordered pair (and the analogous constructions) are low enough that all references to "elements of the set N of natural numbers" can be replaced with references to "elements of elements of the set USC$[N]$ of singletons of natural numbers"; USC$[N]$ is a set in *NFP*. The fatal problem arises in the definition of the map Push from the universe onto the set of functions; we do not know how to show that this map exists in *NFP* (although it can be shown to exist in *NFI*). There may be ways to evade this difficulty, but we have not found one.

3 The Attempt to Interpret *NFP* in *TRCP*

The attempted interpretation of *NFP* in *TRCP* is also analogous to the interpretation of *NF* in *TRC* given in [1], but we give it in more detail. We use the terms π_1 and π_2 to represent the truth-values True and False respectively. We refer to a term f such that "$f(x) = \pi_1$ or $f(x) = \pi_2$ for all x" is a theorem as a "characteristic function term", and use characteristic function terms to represent sets in the natural way.

We represent the logical operations of negation, conjunction, and disjunction by the symbols $\sim$, &, and $\|$, respectively.

We now construct characteristic function terms "$\{x|\phi\}$" for certain formulae ϕ and variables x, which will not contain x and will satisfy $\{x|\phi\}(x) = \pi_1$ if ϕ and $\{x|\phi\}(x) = \pi_2$ if $\sim \phi$. We define $\{x|T = U\}$ as $(\lambda x)(\mathrm{Eq}(T,U))$; we define $\{x| \sim \phi\}$ as $\{x|\{x|\phi\}(x) = \pi_2\}$ and $\{x|\phi\&\psi\}$ as $\{x|(\{x|\phi\}(x), \{x|\psi\}(x)) = (\pi_1, \pi_1)\}$; we define $\{x|(\forall y)(\psi)\}$ as $\{x|\{y|\psi\} = \mathrm{K}[\pi_1]\}$. Of course, these definitions succeed only under certain conditions.

We apply a technique adapted from the proof of the Abstraction Theorem for full *TRC* in [1]: K$[(f,g)]$ = (K$[f]$,K$[g]$) and K$[f(g)]$ = Comp(f,K$[g]$) are easy theorems of *TRCP*, which can be used to eliminate all occurrences of the K-constructor other than iterated applications to atomic terms. This gives us the ability to define abstracts under more general circumstances. If x appears with the same non-negative type n wherever it appears in T and U, and does not appear in the scope of a K-constructor in the simplified form $\mathrm{K}^n[\mathrm{Eq}(T,U)]'$ of $\mathrm{K}^n[Eq(T,U)]$, then it is possible to define $(\lambda x)(\mathrm{Eq}(T,U))$ and thus $\{x|T=U\}$ as $(\lambda x)(\mathrm{Eq}(\mathrm{K}^n[Eq(T,U)]',\mathrm{K}^n[\pi_1]))$. The abstracts used in defining $\{x|\sim\phi\}$ and $\{x|\phi\&\psi\}$ always exist, subject to inductive hypothesis, but this technique may be used to extend the scope of the definition of $\{x|(\forall y)(\psi)\}$. We call a condition ϕ "stratified" if we can assign a type to each variable and a type to each term appearing in the formula (the two sides of an equation are assigned the same type) in such a way that the type of each variable relative to each term in the formula in which it appears is the result of subtracting the type assigned to the term from the type assigned to the variable. We claim that for any stratified condition ϕ in which neither the variable x nor any bound variable nor any K-construct having x or a bound variable in its scope is assigned type higher than that of x (the assigned types can clearly be extended to every subterm of a term in the formula), $\{x|\phi\}$ can be defined. For each subformula $\{x|T=U\}$, the conditions ensure that $(\lambda x)(\mathrm{Eq}(T,U))$ can be defined as indicated in the previous paragraph. If x does not occur in T or U, one must nonetheless modify types as indicated in constructing $(\lambda x)(\mathrm{Eq}(T,U))$ so that the relations between the types of variables in the interpretations of different subformulae is correct (the assigned types of T and U are used to determine the type of the absent occurrences of x). The constructions for formulae constructed by negation and conjunction succeed if they succeed for the subformulae. The construction for $\{x|(\forall y)(\psi)\}$ may run into difficulty if $\{y|\psi\}$ does not exist due to y having too low a type. The trick is that if the type of y is $n-m$, one may convert ψ to a form in which y appears only in the context $\mathrm{K}^m[y]$ (using the simplification above to eliminate complex K-constructs and using (Const) to introduce additional applications of K where necessary). By hypothesis on types of variables and K-constructs, $\mathrm{K}^{m+1}[y]$ will not occur. Replace $\mathrm{K}^m[y]$ by a variable z not found in ψ; the set $\{z|\psi\}$ is defined by inductive hypothesis (induction on length of formulae), and may more instructively be called $\{\mathrm{K}^m[y]|\psi\}$. $\{x|\{y|\psi\}=\mathrm{K}[\pi_1]\}$ may be defined as $\{x|\{\mathrm{K}^m[y]|\psi\}=\mathrm{K}^{m+1}[\pi_1]\}$, where abstraction succeeds. We can then give a "natural" interpretation of "$x\in y$" as "$y(x)=\pi_1\&(\forall z)(y(z)=\pi_1 \parallel y(z)=\pi_2)$". An formula ϕ which translates a formula of *NFP* which is stratified and contains no bound variable with type higher than that of x will satisfy the conditions given above for existence of $\{x|\phi\}$.

We thus successfully obtain (as in [1] for *NF*) an interpretation of the comprehension scheme of *NFP* but without extensionality; each object which is not a "characteristic function" is interpreted as an "ur-element". In [1], we solved this problem by defining a bijection Push from the universe onto characteristic functions and redefining "$x\in y$" as "$\mathrm{Push}(y)(x)=\pi_1$". The interpreted set "$\{x|\phi\}$" is the inverse image under Push of the characteristic function of the collection of objects satisfying the translation of the formula ϕ. We observed that each *TRC* function f can be sent to the function $\mathrm{Setof}(f) = \{(x, y) | f(x) = y\}$. The function Setof exists in *TRCP* as well. We could then define Pushset, the inductive closure of the collection of non-characteristic functions under Setof, and define Push as being Setof on elements of Pushset and the identity elsewhere. The problem here is that (the characteristic function of) Pushset cannot be defined in *TRCP* for the same reason that N cannot be defined in *NFP*; it is defined using a quantifier over all characteristic functions of sets containing the non-characteristic functions and closed under Setof, and these would be assigned the same relative type as that of the set whose characteristic function is being defined. It follows that Push cannot be defined in *TRCP* in the way that it was defined in [1]. Note that the interpretation of *NFP* in *TRCP* or vice versa succeeds in the presence of the local version of the assertion "There is a bijection between the class of functions and the class of characteristic functions"; if there is such a bijection which is predicatively definable, the two theories are equivalent. An assumption sufficient to establish the existence of such a bijection is the Schroder-Bernstein Theorem, which apparently cannot be proven in *NFP*.

4 *TRCI* Introduced

We now introduce the theory *TRCI*. *TRCI* extends *TRCP*. It has an additional term construction: $\mathrm{Abst}[f]$ is a term if f contains no variable of non-negative type and contains no Abst construct except as a subterm of a subterm of negative type; f will be a substitution instance in *TRCI* of a term of *TRCP* containing no variable of non-negative type. If $\mathrm{Abst}[f]$ has relative type n, so does f. The additional axiom scheme of *TRCI* defining the behaviour of Abst is

(Abst′): $\mathrm{Abst}[f](g) = f(\mathrm{K}[h])(g(h))$

The argument for equivalence between *TRCI* and *NFI* succeeds. To see this, we first need the additional

Abstraction Theorem for *TRCI*: Let T be a term of *TRCI* containing no variable of positive type, and let x be a variable which occurs in T with no type other than 0 and does not appear within the scope of any Abst construct or within the scope of any K-construct which itself is within a K-construct. Then $(\lambda x)(T)$ exists as above.

Proof of the Abstraction Theorem for *TRCI*: Use the technique given above to eliminate all complex K-constructs. The only original case of the inductive definition of $(\lambda x)(T)$ which needs to be modified is the case $(\lambda x)(U(V))$, in which it is possible that U may contain $\mathrm{K}[x]$ with type 0 and not within the scope of any K-construct or Abst construct, so that U can be expressed in the form $f(\mathrm{K}[x])$ for some f (using Abstraction for *TRCP*), and $(\lambda x)(U(V))$ can be defined as $\mathrm{Abst}[f]((\lambda x)(V))$ (f will clearly contain no variable of non-negative type); it needs to be observed that $(\lambda x)(\mathrm{K}[A])$ must still be $\mathrm{K}[\mathrm{K}[A]]$, because the only way for it to contain x would be if A were x or an iterated K-construct on x, in which case x would have the wrong relative type. $(\lambda x)(\mathrm{Abst}[U])$ is a trivial case, since U cannot contain x. ∎

Note that the Abstraction Theorem for *TRCI* is *not* stronger than the Abstraction Theorem for *TRCP*, which we will still need to use (with the additional condition that the bound variable not appear in an Abst construct).

5 Equivalence of *TRCI* and *NFI*

We observe that *TRCI* can be interpreted in *NFI*. The point which needs to be established is that the additional functions defined in (Abst′) can be interpreted in *NFI*. The place where *NFI* comprehension is needed is in establishing the existence of the functions $\mathrm{Abst}[f](g)$ for f a constant of *TRCP* and g any function; since $\mathrm{Abst}[f](g)(h)$ is supposed to be $f(\mathrm{K}[h])(g(h))$, a bound variable representing $\mathrm{K}[h]$ will be needed at the same type as that of $\mathrm{Abst}[f](g)$. The dependence of $\mathrm{Abst}[f](g)$ on g is predicative. As in *NFP*, the construction of the Quine ordered pair and related structures presents no difficulty (in fact, no more difficulty in this case than in *NF*), and, as we remarked above, the definition of Push goes through in *NFI* just as in *NF*.

We demonstrate that *TRCI* interprets *NFI*. Let ϕ be a stratified condition in the language of *TRCP* in which no occurrence of a variable or of a K-construct with a variable in its scope has type more than one greater than the type of x. Convert ϕ to a form in which x occurs only in the context $\mathrm{K}[x]$, and define $\{\mathrm{K}[x]|\phi\}$ using the comprehension techniques of *TRCP*–note that $\{x|\phi\}(x) = \{\mathrm{K}[x]|\phi\}(\mathrm{K}[x]) =\{\mathrm{K}[x]|\phi\}(\mathrm{K}[x])(\pi_1,\pi_2)$ can be satisfied by defining $\{x|\phi\}$ as $\mathrm{Abst}[\{\mathrm{K}[x]|\phi\}](\mathrm{K}[(\pi_1,\pi_2)])$ ($\{\mathrm{K}[x]|\phi\}$ cannot contain any variables of non-negative type). This gives us the ability to interpret *NFI* comprehension. The condition "$\mathrm{Push}(y) = z$" can be expressed in the language of *TRCP* using variables one type higher than the type of y: "y belongs to each set A which contains the non-characteristic functions and is closed under Setof and $\mathrm{Setof}(y) = z$, or y fails to belong to some such set and $y = z$". Define "$x \in y$" as "for some z, $z(x) = \pi_1$ and '$\mathrm{Push}(y) = z$' ". Let ϕ be the translation of a stratified, "mildly impredicative" condition on x. The highest type of variable mentioned is two higher than the type of x, occurring in translations of membership statements "$u \in v$" where v is of type one higher than x (and thus a parameter). These can be eliminated by eliminating reference to the parameter v in favor of reference to (the variable representing) $\mathrm{Push}(v)$: i. e., replace each formula "$w(u) = \pi_1$ & '$\mathrm{Push}(v) = w$' " with "$w(u) = \pi_1$", the parameter w replacing the parameter v, wherever v is of the highest possible type; this is legitimate because

"Push(y) $= z$" is provably a bijective relation. Once this elimination is carried out, there are no variables used of type more than one higher than the type of x, and $\{x|\phi\}$ can be defined using the procedure indicated above.

We have established that the theories *TRCI* and *NFI* are precisely equivalent in consistency strength and expressive power.

6 References

1. Holmes, M. Randall. 1991. "Systems of Combinatory Logic Related to Quine's 'New Foundations'". A.P.A.L., 53, pp. 101-33.

2. Quine, W.V. 1937. "New foundations for mathematical logic". American Mathematical Monthly, 44, pp. 70-80.

3. Jensen, R.B. 1969. "On the consistency of a slight (?) modification of Quine's 'New Foundations'". Synthese, 19, pp. 250-63.

4. Crabbé, Marcel. 1982. "On the consistency of an impredicative subsystem of Quine's *NF*". J.S.L., vol. 47, no. 1, pp. 131-6.

5. Holmes, M. Randall. To appear. "Modelling fragments of Quine's 'New Foundations'". To appear in Cahiers du Centre de Logique no. 7, Universite Catholique de Louvain, Institut Superieur de Philosophie, Cabay, Louvain-la-Neuve.

6. Specker, E.P. 1953. "The axiom of choice in Quine's 'New Foundations for Mathematical Logic'". Proc. Natl. Acad. Sci. USA, 39, pp. 972-5.

7. Rosser, J.B. 1978. Logic for Mathematicians, second edition. Chelsea, New York.