**Noname manuscript No.**
(will be inserted by the editor)

# Representation of Functions and Total Antisymmetric Relations in Monadic Third Order Logic

**M. Randall Holmes**

M. Randall Holmes
Department of Mathematics, Boise State University, 1910 University Dr, Boise ID USA 83725 Tel: 1-208-426-3011 E-mail: rholmes@boisestate.edu

### 1 Higher order logics TT and $TT_3$

We start by formalizing higher order logic in order to carefully formulate the question we are addressing.

The theory we present initially is the simply typed theory of sets, equivalently higher order monadic predicate logic of order $\omega$, which we call TT (for "theory of types"). This theory is often confused with the type theory of Russell and Whitehead's [14], but is far simpler: before TT could be formulated, it had to be noted that $n$-ary relations could be implemented as sets via a representation of ordered pair (first done by Wiener in [16]) and the ramifications of the type theory of [14], motivated by predicativist scruples, had to be stripped out, as by Ramsey ([12]). The history of this theory is outlined in [15]: it seems to actually first appear in print about 1930, long after [14]. We are specifically concerned with an initial segment $TT_3$ of this theory.

TT is a first-order theory with sorts indexed by the natural numbers. Its primitive predicates are equality and membership. Atomic sentences $x = y$ are well-formed iff the sorts of the variables $x$ and $y$ are the same. Atomic sentences $x \in y$ are well-formed iff the sort of $y$ is the successor of the sort of $x$. The axiom schemes of TT are extensionality:

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

for each assignment of sorts to $x, y, z$ which yields a well-formed sentence, and comprehension:

$$(\exists A : (\forall x : x \in A \leftrightarrow \phi)),$$

for each formula $\phi$ in which $A$ does not occur free, and for each assignment of sorts to variables which makes sense. The witness to the instance of comprehension associated with a formula $\phi$, which is unique by extensionality, is denoted by $\{x : \phi\}$, a term whose sort is the successor of the sort of $x$.

For each natural number $n$, the theory $\mathrm{TT}_n$ is the subtheory of TT using only the $n$ sorts indexed by $m$ with $0 \leq m < n$. $\mathrm{TT}_n$ is a formalization of $n$th order monadic predicate logic (the logic of unary predicates, that is, properties). Sort 0 is inhabited by individuals; sort $m + 1 < n$ is inhabited by sets of sort $m$ objects representing properties of sort $m$ objects: the axiom of extensionality gives us an identity condition for properties which is defensible though not uncontroversial, and the axiom of comprehension ensures that all properties of a parameter $x$ of sort $m$ which we can represent by a formula of first order logic $\phi(x)$ are in fact represented by sort $m + 1$ objects.

We are interested here in the representation of binary relations and functions in fragments of TT. The existence of the standard Kuratowski pair (for which the index reference is [7]) shows that $\mathrm{TT}_4$ contains a full implementation of second order logic of binary relations on sort 0: a relation represented by a formula $\phi(x,y)$ with sort 0 parameters $x,y$ is represented by $\{\{\{x\},\{x,y\}\} : \phi(x,y)\}$, an object of sort 3.

It is useful to note that there is an internal notion of *finite set* in $\mathrm{TT}_3$. A sort 2 collection $F$ is said to be inductive iff $\emptyset^1 \in A$ and for each $A \in F$ and $x \notin A$, $A \cup \{x\} \in F$. A finite set (of sort 1) is a set belonging to every inductive set (of sort 2).

The precise question that concerns us here is the representability of binary relations and functions in $TT_3$, where the ordered pair of Kuratowski is not available.

It is worth noting that $TT_3$, that is, monadic third order logic, is essentially the logical framework used by David Lewis in his *Parts of Classes* ([8]), so this investigation is relevant to the capabilities of that system.[1] In particular, it is applicable to an inquiry into the extent to which that framework can express quantification over relations. More generally, our investigation fits into a program of justifying logically and mathematically useful concepts with minimal ontological assumptions. It is worth noting in particular that it is known that in the presence of Lewis's framework, various systems of set theory are equivalent to assertions about the cardinality of the universe, which might be thought to give interest to the fact that we investigate definitions of cardinality in monadic third order logic below.

[1] Lewis's framework is articulated in terms of plural quantification and mereology in a way which might make it hard to recognize this. One would interpret sort 2 as inhabited by (singularized) referents of plurally quantified variables, sort 1 as inhabited by fusions of atoms and sort 0 as inhabited by atoms. There are some quibbles about the empty set in either of the sorts of positive index, which admit straightforward resolutions.

**2 Representation of binary relations in $\mathrm{TT}_3$**

To begin with, a fact known from the beginnings of set theory is that reflexive, transitive relations (and so in particular equivalence relations and partial orders) are representable in $\mathrm{TT}_3$. The basic idea is that an order is representable by the collection of its segments. If $x\,R\,y$ represents a formula $\phi(x, y)$ with $x, y$ of sort 0, and this relation is symmetric and transitive in the obvious sense, then $R$ is represented by the set $[R] = \{\{y : y\,R\,x\} : x\,R\,x\}$ of sort 2. The assertion $x\,R\,y$ is equivalent to $y \in \bigcup[R] \wedge (\forall z \in [R] : y \in z \rightarrow x \in z)$. This fact allows us to note that the assertion that there is a linear order on sort 0 can be formulated in $\mathrm{TT}_3$. For any set $A$, we can define a reflexive transitive relation $R_A$ on $\bigcup A$: $x\,R_A\,y$ iff $(\forall z \in A : y \in z \rightarrow x \in z)$. It is the case that $R_{[R_A]}$ is the same relation as $R_A$, though $[R_A]$ will not as a rule be the same set as $A$. Zermelo used this technique to represent well-orderings as sets in his 1908 proof of the Well-Ordering Theorem ([17]): this was important because at that time it was not known how to represent ordered pairs as sets.

Symmetric relations on sort 0 are obviously representable in $\mathrm{TT}_3$ as sort 2 sets of unordered pairs.

If there is a linear order on sort 0 in a model of $\mathrm{TT}_3$ with at least ten individuals (we do not know whether 10 is minimal), then there is a method of defining for sort 0 objects $x, y$ an ordered pair in sort 1, and so all binary relations are representable in sort 2, completely solving the problem of representability of binary relations and functions in $\mathrm{TT}_3$ in this case. Let $\leq$ be a linear order on the universe, represeented internally by the set of its segments as indicated above. Let $a, b, c, d, e, f, g, h, i, j$ be ten distinct sort 0 objects. Define $(x, y)$ as $\{x, y\} \Delta \{a, b, c, d, e\}$ if $x \leq y$ and as $\{x, y\} \Delta \{f, g, h, i, j\}$ otherwise.

The situation described in the previous paragraph can be obtained under a weaker hypothesis. If there is a total antisymmetric relation $C(x, y)$ on sort 0 (a relation $C$ such that $C(x, x)$ is always true, and if $x$ and $y$ are distinct, exactly one of $C(x, y)$ and $C(y, x)$ is true; $C(x, y)$ may be read "$x$ is chosen over $y$") and this relation may be used in instances of comprehension, then a sort 1 ordered pair $(x, y)$ may be defined as $\{x, y\} \Delta \{a, b, c, d, e\}$ if $C(x, y)$ and as $\{x, y\} \Delta \{f, g, h, i, j\}$ otherwise, and all binary relations on sort 0 may be represented as sort 2 sets of ordered pairs in the usual way as in the previous paragraph. If we were in TT or even $\mathrm{TT}_5$, we could understand existence of a total antisymmetric relation as a choice principle, the existence of a choice function from all pairs.

We show that total antisymmetric relations can be represented in $TT_3$ if they satisfy a technical condition weaker than transitivity.

For each $x$, let $C_x$ be defined as $\{y : C(y, x)\}$. Let $C_1$ be defined as
$\{C_x : x = x\}$. Let $C_2$ be defined as $\{C_x \setminus \{x\} : x = x\}$.

We would like to claim that for each $x$, we can define $C_x$ as the unique element $A$ of $C_1$ such that $x \in A$ and $A \setminus \{x\}$ belongs to $C_2$. Certainly $A = C_x$ has this property. Suppose that for some other set $B = C_u \in C_1$, we also have $x \in B$ and $B \setminus \{x\} = C_v \setminus \{v\} \in C_2$. By hypothesis, $A \neq B$, so $x \neq u$. Thus $u \in C_u \setminus \{x\} = C_v \setminus \{v\}$, so $C(u, v)$ and $u \neq v$. We have $C_u = (C_v \setminus \{v\}) \cup \{x\}$. If $v = x$ we would then have $C_u = C_v = C_x$ which we know is false.

So we have a bad case in which there are $u$ and $v$ such that

$$C_u = (C_v \setminus \{v\}) \cup \{x\}$$

and $x \notin C_v$.

This motives the following definition.

**Definition 1** Let $C$ be a total antisymmetric relation on sort 0 of a model of $\mathrm{TT}_3$ understood from context. We define $C_x$ as $\{y : C(y,x)\}$ (as above) for any sort 0 object $x$, For any sort 0 object $x$, a pair $\{u,v\}$ is called a *bad pair (in $C$) with respect to $x$* if we have $u \neq v$, $u\,C\,v$, $x \notin C_v$, and $C_u = (C_v \setminus \{v\}) \cup \{x\}$. We summarize some consequences: this gives us $\neg x\,C\,v$, so $v\,C\,x$, $x\,C\,u$, so $\neg u\,C\,x$. All $w$ not in $\{u,v,x\}$ satisfy $w\,C\,u \leftrightarrow w\,C\,v$. A pair $\{u,v\}$ is simply called a *bad pair* iff there is an $x$ such that $\{u,v\}$ is a bad pair with respect to $x$.

We can then rule out this bad case by modifying our attempt to define $C_x$: $C_x$ is the unique element $A$ of $C_1$ such that $x \in A$ and $A \setminus \{x\} \in C_2$, and further there is no $B \in C_1$ and $v$ of sort 0 such that $x \notin B$ and $(B \setminus \{v\}) \cup \{x\} = A$. The additional condition rules out the alternative possibility that $A = C_u$ where $\{u, v\}$ is a bad pair for $x$.

With the new definition, the only way that $C_x$ can fail to be defined is if there is a bad pair $\{u, v\}$ with respect to $x$ and $x$ itself is a member of a bad pair $\{x, s\}$ with respect to some $t$. Note that if $\{u, v\}$ is a bad pair, $u$ and $v$ have the same $C$ relations to every object other than $u, v, x$. Thus if there is a bad pair $\{u, v\}$ with respect to $x$ and $x$ itself is a member of a bad pair $\{x, s\}$ with respect to some $t$, we have that either $s$ is one of $u, v$ or that $x$ and $s$ have the same $C$ relations to $u, v$, and the latter is impossible, since this would mean that $s$ had different $C$ relations to $u, v$. If $s = u$, we know that $x \in C_u$, so it is $\{x, u\}$ that is the bad pair, and $C_x = C_u \setminus \{u\} \cup \{t\}$. The only thing that $t$ can be is $v$, as we know that $v \in C_x$ (as $x \notin C_v$) and $v \notin C_u$. It further follows that $C_x \setminus \{x\} \cup \{u\} = (C_u \setminus \{u\} \cup \{v\}) \setminus \{x\} \cup \{u\} = C_u \setminus \{x\} \cup \{v\} = C_v$, so $\{v, x\}$ is also a bad pair. Similar reasoning shows that if $s = v$ we also have $\{v, x\}$ and $\{x, u\}$ bad pairs with respect to $u$ and $v$ respectively.

This motivates a definition.

**Definition 2** Let $C$ be a total antisymmetric relation on sort 0 of a model of $\mathrm{TT}_3$ given in the context. For any $u, v, x$ of sort 0, we say that $\{u, v, x\}$ is a *bad triple (in $C$)* iff $\{u, v\}$ is a bad pair in $C$ with respect to $x$ and $\{v, x\}$ is a bad pair in $C$ with respect to $u$ and $\{x, u\}$ is a bad pair in $C$ with respect to $v$. The discussion above implies that any two of these conditions imply the third. Note that we have $u\,C\,v$, $v\,C\,x$ and $x\,C\,u$ and further that for any $w \notin \{u, v, x\}$ we must have that $w\,C\,x$, $w\,C\,u$, $w\,C\,v$ all have the same truth value (from which it follows that distinct bad triples must be disjoint). We further define the *circulation* of $C$ as the relation $C^\circ$ such that $x\,C^\circ\,y$ holds iff either $x = y$ and $x$ is not an element of any bad triple in $C$, or $x, y$ are two of the elements of some bad triple in $C$ and $x\,C\,y$. Note that the circulation of $C$ is in every case a function (in fact a bijection), permuting the elements of each bad triple and fixing all other sort 0 objects.

Thus we can assert the existence of a particular kind of total antisymmetric relation (one which has no bad triples) in the language of $\mathrm{TT}_3$ by asserting the existence of sets $D$ and $E$ such that for each $x$ of sort 0 there is a unique $D_x \in D$ such that $x \in D_x$ and $D_x \setminus \{x\} \in E$, and no $B \in D$ and $v$ satisfy $x \notin B$ and $D_x = (B \setminus \{v\}) \cup \{x\}$, and satisfying the additional condition that for each $x$ and $y$ distinct, exactly one of $x \in D_y$ and $y \in D_x$ holds: one can then define $C(x, y)$, a total antisymmetric relation, as $x \in D_y$, and define an ordered pair of sort 0 objects in sort 1 and so a complete representation of binary relations on sort 0 in sort 2 as above. The technical condition on the relation that it has no bad triples follows from the claimed conditions on $D$ and $E$ as above; it does not need to appear in the claimed conditions. That the existence of a total antisymmetric relation with no bad triples implies the existence of such sets representing it is shown above.

### 3 Representation of a large class of functions in $\mathbf{TT_3}$

In the absence of any choice principles, we present a result about representability of a wide class of functions. We state to begin with that we will focus on representing functions taking sort 0 objects to sort 0 objects which are of universal domain (defined on all of sort 0). When we do want to represent partial functions with a given domain, each function $f$ with domain $D$ a proper subset of sort 0 will be identified with the extension of $f$ which agrees with $f$ on $D$ and acts as the identity function on the complement of $D$.

**Definition 3** We fix a sort 0 variable $x$ and a sort 0 variable $y$. We call a formula $\phi$ *functional* iff $(\forall x : (\exists y : \phi) \wedge (\forall xyz : \phi \wedge \phi[z/y] \rightarrow y = z)$ holds. When $\phi$ is functional, we will usually write $\phi(u, v)$ for $\phi[u/x][v/y]$, the result of substituting $u$ for $x$ and $v$ for $y$ in $\phi$, so the condition already stated can be written

$$(\forall x : (\exists y : \phi(x, y)) \wedge (\forall xyz : \phi(x, y) \wedge \phi(x, z) \rightarrow y = z).$$

We write $f_\phi(x)$ for the unique $y$ such that $\phi(x, y)$. For any set $A$, we let $f_\phi\lceil A$ abbreviate $f_{(x \in A \wedge \phi) \vee (x \notin A \wedge y = x)}$ [this is an example of the treatment of partial functions announced above].

**Definition 4** If $\phi$ is a functional formula and $A$ is a sort 1 set, we say that $A$ is closed under $f_\phi$ iff $(\forall x \in A : \phi(x,y) \rightarrow y \in A)$. If $x \in \mathtt{dom}(\phi)$ we define $\mathtt{orbit}_\phi(x)$, the forward orbit of $x$ in $f_\phi$, as the intersection of all sets which are closed under $f_\phi$ and contain $x$ as an element. We define a finite cycle in $f_\phi$ as a finite set $\mathtt{orbit}_\phi(x)$ such that for each $y \in \mathtt{orbit}_\phi(x)$, $\mathtt{orbit}_\phi(x) = \mathtt{orbit}_\phi(y)$. We are interested in finite cycles of cardinality greater than two: by this we simply mean finite cycles which are not singletons or unordered pairs (we do not presuppose a development of the notion of cardinality by using this phrase).

**Theorem 1** *We work in an arbitrary model of $TT_3$. There is a uniform way to represent functional formulas $\phi$ by sets $[f_\phi]$ for each $\phi$ for which there is a choice set $C_\phi$ for finite cycles in $f_\phi$ of cardinality greater than 2.*

*Proof* The set $[f_\phi]$ which we take as representing the function $f_\phi$ is the set of all items of the following kinds:

1. forward orbits in the restriction $f_\phi\lceil(V^1 \setminus C_\phi)$. ($V^1$ being the sort 1 set of all sort 0 objects). It is important to note that in accordance with our convention about partial functions, $f_\phi\lceil(V^1 \setminus C_\phi)$ fixes each element of $C_\phi$. It is also important to note that every forward orbit in $f_\phi$ is also a forward orbit of this restriction.
2. singletons of elements of $C_\phi$
3. singletons of elements of $f_\phi$“$C_\phi = \{y : (\exists x \in C_\phi : \phi(x, y)))\}$.

Given a set $F$, we indicate how to reverse engineer a functional formula $\phi$ such that $F = [f_\phi]$ if there is one, and how to recognize when there is no such formula.

Note first that if $F = [f_\phi]$, then $\bigcup F = V^1$.

Notice next that in any function representation $F = [f_\phi]$, an element $A$ includes a finite cycle in $f_\phi$ of cardinality $> 2$ as a subset if and only if it includes exactly two singletons belonging to $F$ as subsets. The element $A$ is a finite cycle in $f_\phi$ of cardinality $> 2$ iff it has the previous property and in addition no proper subset of $A$ which belongs to $F$ includes two singletons belonging to $F$ as subsets. Further, if $A$ is a finite cycle in $f_\phi$, each of its proper subsets which belongs to $F$ and is not a singleton will include the singleton of the element of $A$ which belongs to $C_\phi$ as a subset and no proper subset of $A$ which belongs to $F$ and is not a singleton will include the singleton of the element of $A$ which belongs to $f_\phi$"$C_\phi$ as a subset.

This motivates the following

**Definition 5** Let $F$ be an arbitary sort 2 set such that $\bigcup F = V^1$. The collection of *supercycles of* $F$ is defined as the collection of all elements of $F$ which include exactly two singletons belonging to $F$ as subsets. The collection of *cycles of* $F$ is defined as the collection of all supercycles of $F$ which have no proper subsets which are supercycles of $F$. We define $C_F$ as the collection of all $x$ such that $\{x\} \in F$ and for some cycle $A$ in $F$ of cardinality $> 2$, $x \in A$ and every proper subset $B \in F$ of $A$ has $x$ as an element. We define $D_F$ as the collection of all $x$ such that $\{x\} \in F$ and for some cycle $A$ in $F$, $x \in A$ and $\{x\}$ is disjoint from each proper subset of $A$ belonging to $F$ other than $\{x\}$. We say that $F$ is $C$-good if $\bigcup F = V^1$ and each cycle of $F$ is finite and contains as elements exactly one element of $C_F$ and exactly one element of $D_F$.

Further note if $F = [f_\phi]$, the forward orbits in $f_\phi$ are exactly those sets which are either supercycles in $F$ or not included in any supercycle in $F$. The forward orbit of any sort 0 object $x$ is the intersection of all forward orbits containing $x$. Further, the forward orbits in $f_\phi\lceil(V^1 \setminus C_\phi)$ are exactly those elements of $F$ which are not singletons of elements of $D_F$. This motivates the following

**Definition 6** Let $F$ be any $C$-good sort 2 set. Define $F^*$ as the set of all elements of $F$ which are either supercycles of $F$ or not included in any supercycle of $F$. For any sort 0 object $x$, define $\mathtt{Orbit}_F(x)$ as the intersection of all elements of $F^*$ which contain $x$. Define $F^{**}$ as the set of all elements of $F$ which are not singletons of elements of $D_F$. Define $\mathtt{Orbit}^*_F(x)$ as the intersection of all elements of $F^{**}$ which contain $x$. We say that a $C$-good set $F$ is orbit-good iff each $\mathtt{Orbit}_F(x)$ is an element of $F$, each $\mathtt{Orbit}^*_F(x)$ is an element of $F$, and all elements of $F$ are either $\mathtt{Orbit}_F(x)$'s, $\mathtt{Orbit}^*_F(x)$'s, singletons of elements of $C_F$ or singletons of elements of $D_F$.

Further, note that for any element $x$ of $V^1 \setminus C_\phi$, $f_\phi(x)$ is the unique $y$ in the forward orbit $O$ of $x$ in $f_\phi\lceil(V^1 \setminus C_\phi)$ such that the forward orbit of $y$ in $f_\phi\lceil(V^1 \setminus C_\phi)$ is either $O \setminus \{x\}$, or is equal to $O$ which is equal to $\{x, y\}$ (this last case does not exclude the possibility that $x = y$). For each element $x$ of $C_\phi$, $f_\phi(x)$ is the element of $f_\phi``C_\phi$ contained in the same finite cycle in $f_\phi$. This motivates the following

**Definition 7** For any orbit-good $F$ and $x$ of sort 0, we define $F[x]$ as follows:

1. If $x$ belongs to $C_F$, define $F[x]$ as the element of $D_F$ belonging to the same cycle in $F$.
2. If $x$ does not belong to $C_F$, define $F[x]$ as the unique $y$ such that either $\mathtt{Orbit}^*_F(y) = \mathtt{Orbit}^*_F(x) \setminus \{x\}$ or $\mathtt{Orbit}^*_F(y) = \mathtt{Orbit}^*_F(x) = \{x, y\}$ (which does not rule out $y = x$, note).

We say that $F$ is value-good iff $F$ is orbit-good, $F[x]$ is defined for every $x$ and further for each $x$ the minimal set $O(x)$ such that $x \in O(x)$ and $(\forall y : y \in O(x) \to F[y] \in O(x))$ satisfies $O(x) = \mathtt{Orbit}_F(x)$.[2]

[2] An example of a value-good $F$ which would not be orbit-good would be the collection of final segments of an infinite well-ordering with order type $> \omega$).

We have now described precisely how to determine for any $F$ whether it represents a function and what the extension of the represented function is. The value-good sets are the sets which represent functions, and for each value-good $F$ we have $F = [f_{y=F[x]}]$, where of course $y = F[x]$ abbreviates a very complicated formula.

Notice that under the hypothesis $\mathrm{AC}_{\mathtt{fin}}$ = "every collection of pairwise disjoint finite sets has a choice set", every function is representable in this sense.

#### 4 Applications: cardinality can be represented in $\mathrm{TT}_3$ and $\mathrm{NF}_3$; more about total antisymmetric functions

An immediate application of this partial representation of functions is a demonstration that the notion of cardinality is definable in $\mathrm{TT}_3$ (for sets of sort 1). It is not the case that every bijection is representable in this way. However, if there is a bijection $f_\phi$ from a set $A$ to a set $B$ which is represented by a formula $\phi(x, y)$ as discussed above (extended to act as the identity function on non-elements of $A$), there is also a representable function $f^*$ whose restriction to $A$ is a bijection from $A$ to $B$ and which acts outside $A$ as the identity. The value $f^*(x)$ for $x \in A$ is defined as $x$ if $x$ belongs to a finite cycle of cardinality greater than 2 in $f_\phi$ (which will be a subset of $A \cap B$) and otherwise as $f_\phi(x)$. The function $f^*$ is clearly both representable by a formula and representable by a set $[f^*]$ defined as above. An application of this is the observation that the notion of cardinality is definable in the fragment $\mathrm{NF}_3$ of Quine's New Foundations (the set theory described in [11], usually abbreviated NF) shown to be consistent by Grishin ([4]). This was shown by somewhat different methods in unpublished work by Henrard (discussed in [9], [3]). That cardinality is definable in $\mathrm{NF}_3$ is not obvious, as there is no notion of ordered pair definable in this theory. It is elegant that the notion of cardinality that we are able to define is such that the domain and range of any bijective functional relation defined by a formula will be of the same cardinality, even if we cannot represent the function by a set. Since we have defined the notion of sets $A$ and $B$ (of sort 1 in $\mathrm{TT}_3$) having the same cardinality, we do have the ability to define the cardinal $|A|$ as the (sort 2 in $TT_3$) collection of all sets $B$ which are of the same cardinality as $A$.

We regard it as worth noting that considerations about $NF_3$ are actually very general considerations about third order logic. We outline the reasons for this. NF can briefly be described as the one-sorted first order theory with equality and membership whose axioms are the axioms of TT with distinctions of sort between variables dropped (without creating identifications between variables); $NF_n$ has the same relationship to $TT_n$. $NF_4$ was shown in [4] to be the same theory as NF. Any two externally infinite models of $TT_2$ with the splitting property (any set which is externally infinite can be partitioned into two externally infinite sets) which have the same cardinality are isomorphic by a back-and-forth construction. Any model of $TT_3$ which is externally infinite[3] is readily shown to be elementarily equivalent to a countable model of $TT_3$ which is externally infinite and has the splitting property. A countable model of $TT_3$ which is externally infinite and has the splitting property possesses an isomorphism from the substructure consising of sorts 0 and 1 to the substructure consisting of sorts 1 and 2, by the observation about $TT_2$ above, and so can be made into a model of $NF_3$ by using the isomorphism to identify the sorts, by results of Specker in [13]. The net effect of this is that the stratified theorems of $NF_3$ (the ones which can be read as theorems of $TT_3$ by assigning sorts to variables) are in fact the theorems which hold in all externally infinite models of $TT_3$ (including externally infinite models of

[3] By "externally infinite" we simply mean that the model is infinite in terms of the metatheory. We emphasize this because an externally infinite model of $TT_3$ may satisfy the negation of the Axiom of Infinity: it may believe internally that all sets are finite.

$TT_3$ in which the axiom of infinity is false): $NF_3$ is in effect a very general system of third order logic. The original reference for this fact is [1]. $NF_4$, on the other hand can be viewed as a very odd system of fourth order logic, and NF can be viewed as a similarly odd system of higher order logic of order $\omega$. It is well-known that NF is strange and presents vexed problems: the point of this paragraph is that $NF_3$, though perhaps unfamiliar to the reader, is not particularly strange and in fact is rather generic. The results of this paper show something about the mathematical competence of this system. It is worth mentioning the result of Pabion ([10]) that $NF_3$ with the Axiom of Infinity is equivalent in strength to second order arithmetic.

Another application of the partial representation of functions is a stronger representation of total antisymmetric relations: let $C$ be a total antisymmetric relation such that there is a choice set from its bad triples: represent $C$ by three sets, $C_1$ defined as above, $C_2$ defined as above, and $C_3$ the set representing the circulation $C^\circ$ of $C$ (defined above) as a function in the way just described, using the given choice set to handle its nontrivial cycles, the bad triples of $C$. $C_x$ can then be defined as in the partial represention of total antisymmetric relations given above, when $x$ does not participate in a bad triple: when $x$ belongs to a bad triple, $C_3$ provides the needed additional information.

The condition asserting the existence of such a representation of a total antisymmetric relation follows: there are sets $D$, $E$, and $F$ such that for each $x$ we are given either a unique $D_x \in D$ such that $x \in D_x$ and $D_x \setminus \{x\} \in E$, and no $B \in D$ and $v$ satisfy $x \notin B$ and

$D_x = (B \setminus \{v\}) \cup \{x\}$, or a unique set $D_x^-$ and pair of sort 0 objects $u, v$ distinct from $x$ and from each other such that each union of $D_x^-$ with a two element subset of $\{x, u, v\}$ belongs to $D$ and each union of $D_x^-$ with a one-element subset of $\{x, u, v\}$ belongs to $E$. We refer to $\{x, u, v\}$ as a bad triple in the latter case. The additional conditions are asserted to hold that for any distinct $x, y$, exactly one of $x \in D_y$ and $y \in D_x$ holds (if $D_x$ and $D_y$ are defined): if $D_x$ and/or $D_y$ are not defined, the same statement holds with $D_x^-$ and/or $D_y^-$ in place of $D_x$ and/or $D_y$, respectively, if it is not the case that both $D_x^-$ and $D_y^-$ are both defined and are the same set (that is, if $x$ and $y$ do not belong to the same bad triple). $F$ is the set representation of a bijection whose domain is the union of the bad triples and which has each bad triple as a cycle. The relation $C(x, y)$ is defined as "$x \in D_y \vee x \in D_y^- \vee y = F[x] \vee y = x$". Of course, if this condition holds we can define a complete implementation of binary relations.

Note that under the hypothesis $\mathrm{AC}_3$ = "every pairwise disjoint collection of three-element sets has a choice set", any total antisymmetric relation has such a set implementation, and we can express in the language of $\mathrm{TT}_3$ the assertion that there is a total antisymmetric relation. We are not saying that $\mathrm{AC}_3$ implies that there is such a relation; we see no reason to believe this to be true.

## 5 There is no uniform representation of functions or of total antisymmetric relations in $\mathrm{TT}_3$

We now present the negative result that there is no uniform way in which all functions representable by functional formulas can be represented by sets in $\mathrm{TT}_3$, nor is there any uniform way to represent total antisymmetric relations representable by formulas as sets. First we state precisely what we mean.

**Definition 8** We say that a *formal implementation of functions in $TT_3$* is constituted by two formulas $\mathtt{fun}_F$ and $\mathtt{app}$ satisfying conditions which we describe. $\mathtt{fun}_F$ is a formula in a language extending the language of $\mathrm{TT}_3$ with a new primitive binary function symbol $F(x, y)$ for a binary relation with parameters of sort 0. The variable $f$ (of a sort we choose not to specify) is the only variable free in $\mathtt{fun}_F$: we will usually write it $\mathtt{fun}_F(f)$ in order to signal this. $\mathtt{app}$ is a formula in the language of $\mathrm{TT}_3$ without $F$ in which the sort 0 variables $x$ and $y$ and the same variable $f$ of sort not stated are the only free variables: we will usually write $\mathtt{app}(f, x, y)$ to signal this. In the extension of $\mathrm{TT}_3$ with the addition of axioms that $F(x, y)$ is a functional formula and that all instances of the comprehension scheme for $\mathrm{TT}_3$ involving the new primitive relation $F$ hold, with no other additional axioms, we require that $(\exists f : \mathtt{fun}_F(f))$ is a theorem and that $\mathtt{fun}_F(f) \rightarrow (\mathtt{app}(f, x, y) \leftrightarrow F(x, y))$ is a theorem.[4]

**Definition 9** Similarly, we say that a *formal implementation of total antisymmetric relations in $TT_3$* is constituted by two formulas $\mathtt{tarel}_R$ and $\mathtt{tarelapp}$ satisfying conditions which we describe. $\mathtt{tarel}_R$ is a formula in a language extending the language of $\mathrm{TT}_3$ with

[4] The single variable $f$ may be replaced throughout by a finite vector $f_1, \ldots, f_n$, if the representation uses more than one object: for example, the representation of functions we would obtain if we had a total antisymmetric relation would consist of the usual collection of ordered pairs representing the function, but also the three sets coding the total antisymmetric relation and the ten sort 0 objects used in the definition of the ordered pair.

a new primitive binary function symbol $R(x, y)$ for a binary relation with parameters of sort 0. The variable $r$ (of a sort we choose not to specify) is the only variable free in $\texttt{tarel}_R$: we will usually write it $\texttt{tarel}_R(r)$ in order to signal this. $\texttt{tarelapp}$ is a formula in the language of $\mathrm{TT}_3$ without $R$ in which the sort 0 variables $x$ and $y$ and the same variable $r$ of sort not stated are the only free variables: we will usually write $\texttt{tarelapp}(r, x, y)$ to signal this. In the extension of $\mathrm{TT}_3$ with the addition of axioms that $R(x, y)$ is a total antisymmetric relation and that all instances of the comprehension scheme for $\mathrm{TT}_3$ involving the new primitive relation $R$ hold, with no other additional axioms, we require that $(\exists r : \texttt{tarel}_R(r))$ is a theorem and that $\texttt{tarel}_R(r) \rightarrow (\texttt{tarelapp}(r, x, y) \leftrightarrow R(x, y))$ is a theorem.[5]

We leave it to the reader to evaluate our assertion that this formalizes exactly what we mean by saying that there is a uniform implementation of functions or of total antisymmetric relations as sets in $\mathrm{TT}_3$. The intended sense of $\texttt{fun}_F(f)$ is "$f$ is the set implementation of the functional binary relation $F$"; the intended sense of $\texttt{app}(f, x, y)$ is "$y$ is the result of applying the function represented by the set $f$ to $x$". The intended sense of $\texttt{tarel}_R(r)$ is "$r$ is the set implementation of the total antisymmetric relation $R$"; the intended sense of $\texttt{tarelapp}(r, x, y)$ is "$x \, R \, y$, where $R$ is the total antisymmetric relation represented by $r$".

We use a Fraenkel-Mostowski permutation model to demonstrate our negative result (a textbook reference for this method is [6]). At this point we stipulate that

[5] As above, the single variable $r$ may be replaced with a finite vector of variables $r_1, \ldots, r_n$: for example, the partial representation of total antisymmetric relations already given has three components.

our metatheory is ZFA (the usual set theory ZFC with extensionality weakened to allow atoms) and that we assume the existence of infinitely many atoms. It is well-known that ZFA with a collection of atoms of any desired size is mutually interpretable with ZFC. A much weaker metatheory could be used, but this one is conventional.

We also note that any model of $\mathrm{TT}_3$ in which the set implementing sort 0 is not larger than the collection of atoms is isomorphic to a model of $\mathrm{TT}_3$ in which sort 0 is implemented by a set of atoms, sort 1 is implemented by a subset of the power set of the set implementing sort 0, sort 2 is implemented by a subset of the power set of the set implementing sort 1, and the membership relations of the model are subrelations of the membership relation of the metatheory. We call such a model of $\mathrm{TT}_3$ a "natural model" of $\mathrm{TT}_3$ in ZFA. Natural models of $\mathrm{TT}_n$ for any finite $n$ can be defined similarly.

**Theorem 2** *There is no formal implementation of functions in $TT_3$, nor is there any formal representation of total antisymmetric relations in $TT_3$.*

*Proof* We set out to construct a natural model of $\mathrm{TT}_4$ in ZFA in which the set of atoms implementing sort 0 is infinite and partitioned into three element sets, which are orbits under a bijection $f$ from sort 0 to sort 0 in the metatheory. We add a new predicate $F(x, y)$ to our language, with the meaning $y = f(x)$. We will allow the predicate $F$ to be used in instances of comprehension. We use the convention that any permutation $\pi$ of the atoms is extended to all sets by the rule $\pi(A) = \pi``A$. The group $G$ of permutations defining the FM model will be the permutations of sort 0 which act on each orbit in $f$ independently as either the identity, $f$ or $f^2 = f^{-1}$. A set or atom $A$ is said to be symmetric iff there is a finite set $S$ of atoms such that for any permutation $\pi \in G$ such that $\pi(s) = s$ for each $s \in S$, we also have $\pi(A) = A$: it is obvious that each atom is symmetric. A set belongs to the FM model iff it is hereditarily symmetric in this sense; all atoms belong to the FM model. Standard results about FM models tell us that we obtain an interpretation of ZFA (without Choice) in our original ZFA in this way. Sort 0 of our model of $\mathrm{TT}_4$ will consist of the set of atoms already mentioned. Sort 1 of our model of $\mathrm{TT}_4$ will be the power set of the set implementing sort 0 in the sense of the FM interpretation. Sort 2 of our model of $\mathrm{TT}_4$ will be the power set of the set implementing sort 1 in the sense of the FM interpretation. Sort 3 of our model of $\mathrm{TT}_4$ will be the power set of the set implementing sort 2 in the sense of the FM interpretation. This is clearly a model of $\mathrm{TT}_4$ both in the FM interpretation and in our original ZFA metatheory, also satisfying the assertion that $F(x, y)$ is a functional formula and satisfying all instances of comprehension mentioning $F$: we can see this because the usual Kuratowski implementation of $f$ is a set in the model of $\mathrm{TT}_4$.

A set of sort 1 in this model is of the form $S \cup T$ where $S$ is a finite set and $T$ is a union of orbits in $f$. The closure of $S$ under $f$ is a support of this set. A set of sort 2 with support $S$, a finite set closed under $f$, is an arbitrary union of basis sets, each one determined by a finite subset $A$ of $S$ and a function $g$ from the orbits of $f$ not included in $S$ to $\{0, 1, 2, 3\}$ which has only finitely many domain elements mapped to 1 or 2. The basis element determined by $A$ and $g$ is the collection of all sets $A \cup B$ where $B$ does not meet $S$ and for each orbit $o$ in $f$ which does not meet $S$ we have $|B \cap o| = g(o)$. These descriptions of sort 1 and sort 2 sets follow directly from the criterion that a set of a given sort in our model with a given support $S$ is an arbitrary union of orbits of permutations in our group which fix the given support.

Now observe (it is evident from the descriptions of sort 1 and sort 2 sets) that the model of $\mathrm{TT}_3$ consisting of sorts 0,1,2 of the model of $\mathrm{TT}_4$ which we have constructed has the property that all of its sets are hereditarily symmetric with respect to the larger group $G^*$ of permutations which fix each orbit of $f$ and act within each orbit entirely arbitarily. But it is still the case that all instances of comprehension mentioning $F$ hold in this model: this property is inherited from the model of $\mathrm{TT}_4$ defined with the smaller group $G$.

By examination of the model of $\mathrm{TT}_3$ just described as an initial segment of the model of $\mathrm{TT}_4$ we started with, we can show that in fact there can be no formal implementation of functions as sets. For if there were such an implementation based on given formulas $\mathtt{fun}_F$ and $\mathtt{app}$, we would be able to identify $f$ such that $\mathtt{fun}_F(f)$ (letting $F$ denote the specific functional relation we introduced in the model construction). Now the object $f$ would have to have a finite support set $S$: for any permutation $\pi \in G^*$ fixing each element of this finite set $S$, we would have $\pi(f) = f$.

It is straightforward to show that for any permutation $\pi \in G^*$ we will have $\mathtt{app}(f, x, y) \leftrightarrow \mathtt{app}(\pi(f), \pi(x), \pi(y))$. This follows from the fact that each atomic formula $u = v$ or $u \in v$ ($F$ will not appear in $\mathtt{app}$) is invariant under application of any $\pi \in G^*$ to both sides, and induction on the structure of formulas. And this cannot be true. Choose any $x, y$ which are not in $S$ such that $y = f(x)$ and choose $\pi \in G^*$ such that $\pi(y) = f^{-1}(\pi(z))$ (we can do this because each orbit in $F$ can be permuted in any arbitrary way by elements of $G^*$), and this falsifies the theorem relating $\mathtt{app}$ and $\mathtt{fun}_F$: we would have $\mathtt{fun}_F(f) \wedge \mathtt{app}(f, x, y)$, from which we could deduce $\mathtt{fun}_F(\pi(f)) \wedge \mathtt{app}(\pi(f), \pi(x), \pi(y))$ (noting that we have $\pi(f) = f$), from which we have both $F(\pi(x), \pi(y))$, by the fact that this is supposed to be a formal representation of functions, and $F(\pi(y), \pi(x))$ by the choice of $\pi$, which is impossible.[6]

We can further show that there can be no representation of total antisymmetric relations in the same sense. The exact model we are considering supports a total antisymmetric relation (representable in the usual way as a set of sort 3). There is a linear ordering $\leq$ of the orbits under $f$ because we are in ZFA with Choice. The total antisymmetric relation defined by "the orbit of $x$ in $f$ is strictly less than the orbit of $y$ in $f$ or $y = f(x)$ or $y = x$" is invariant under permutations in $G$ and so is present in the FM interpretation. If we add a primitive predicate representing this relation, all instances of comprehension mentioning this predicate will hold in the model of $\mathrm{TT}_4$ and in the model of $\mathrm{TT}_3$ which is its initial segment. No formulas $\mathtt{tarel}_R(r)$ and $\mathtt{tarelapp}(R, x, y)$ in the language of $\mathrm{TT}_3$ (in the first formula augmented with a total antisymmetric relation $R$) can constitute a formal representation of total antisymmetric relations by a very similar argument to

[6] Note that the argument goes in exactly the same way if the single variable $f$ representing the function is replaced by a finite vector $f_1 \ldots, f_n$.

that given above. Let $r$ satisfy $\mathtt{tarel}_R(r)$ where $R$ denotes the total antisymmetric relation defined above in terms of $f$. Let $S$ be a support of $r$ with respect to $G^*$. Let $x, y$ be chosen such that neither belongs to $S$ and $y = f(x)$. Let $\pi \in G^*$ fix each element of the support $S$ and satisfy $\pi(y) = f^{-1}(\pi(x))$. We would have $\mathtt{tarel}_R(r) \wedge \mathtt{tarelapp}(r, x, y)$, from which we could deduce $\mathtt{tarel}_R(\pi(r)) \wedge \mathtt{tarelapp}(\pi(r), \pi(x), \pi(y))$ (noting that we have $\pi(f) = f$, and that $\mathtt{tarelapp}(r, x, y) \leftrightarrow \mathtt{tarelapp}(\pi(r), \pi(x), \pi(y))$ for reasons already discussed in connection with $\mathtt{app}$), from which we have both $R(\pi(x), \pi(y))$, by the fact that this is supposed to be a formal representation of functions, and $R(\pi(y), \pi(x))$ by the choice of $\pi$, which is impossible.[7]

This has a corollary with an ironic flavor: if we provide a predicate $R$ representing the total antisymmetric relation described above, we do obtain an ordered pair on sort 0 in sort 1 and a representation of binary relations and so of functions in this model: this does not contradict our results here because the definition of ordered pair and so the definition of a relation holding between two objects or application of a function to an object depend essentially on $R$. This unintended representation of relations and functions can be killed by allowing permutations in $G$ to exchange orbits as well as permute objects independently in each orbit. We do not know whether we can express the assertion that there is some total antisymmetric relation on sort 0 in the language of $\mathrm{TT}_3$: we have shown above that we can express the assertion that there is some total antisymmetric relation on sort 0 if we have the additional hypothesis

[7] Note that the argument goes in exactly the same way if the single variable $r$ representing the total antisymmetric relation is replaced by a vector $r_1, \ldots, r_n$.

$AC_3$ that each collection of disjoint three-element sets has a choice set.

We make a final (cautionary) remark about choice principles in these truncated models of type theory. The choice principle $AC_2$ = "every disjoint collection of pairs has a choice set" holds in the model of $TT_3$ under consideration, because all hereditarily symmetric pairwise disjoint collections of sort 1 (unordered) pairs of sort 0 objects are finite. However, if we build a model of $TT_5$ in the same way in the FM interpretation using $G^*$, we will find that $AC_2$ fails for sort 2 objects: the existence of a choice function for pairs can be proved from the existence of a choice set for the collection of unordered pairs of the form $\{(x, y), (y, x)\}$ where $x, y$ are of sort 0 and the ordered pairs are Kuratowski pairs, and it is straightforward to argue that the FM interpretation using $G^*$ cannot enjoy a choice function for pairs.

This shows that the result on representability of functions above is something like the best possible: the limitation that one must be able to choose an element from each finite cycle of length greater than two has something to do with actual obstructions that can prevent representability of functions in the absence of choice.

It is also worth noting the corollary of the negative result that there is no ordered pair of sort 0 objects definable in sort 1 in $TT_3$ without additional hypotheses, as otherwise there would clearly be a formal representation of functions as sets along standard lines.

### 6 Related work

We have already noted the unpublished work of Henrard on the definability of cardinality in $NF_3$, which was the original inspiration of this work. The only accessible sources known to us which discuss this work are the master's theses [9], [3]. Henrard's aim was to represent cardinality, not functions per se, in the theory $NF_3$ in which no ordered pair is available. He represented orbits in a bijection $f$ as sets of pairs $\{x, f(x)\}$: an orbit would be a minimal set of such pairs closed under the relation of having nonempty intersection, in which each pair $\{x, y\}$ intersected no more than two pairs $\{u, x\}$ and $\{y, v\}$ (and might intersect one pair or none). Notice that the representations of the orbits of $f$ and $f^{-1}$ are indistinguishable. It is then reasonably straightforward to give a definition of the conditions under which a set of pairs would be the union of the representations of the orbits in a bijection from a set $A$ to a set $B$, thus allowing the definition of the notion of sets $A$ and $B$ having the same cardinality, though without actually providing a formal representation of a bijection from $A$ to $B$: we do not give the details. Our approach was developed with prior knowledge of his, and betters it by providing an actual representation of some bijection from $A$ to $B$ when there is any bijection from $A$ to $B$ (though not of all such bijections), and providing representations of many functions which are not bijections. Our results give more information about the mathematical competence of $TT_3$ and $NF_3$ than Henrard's methods: we acknowledge that we are indebted to his work. We believe that it is important to note (as we do at length above) that $NF_3$ is not a special case: every externally infinite model of $\mathrm{TT}_3$ is elementarily equivalent to a model of $\mathrm{NF}_3$ (in the sense that the stratified assertions true in the model of $\mathrm{NF}_3$ correspond exactly to the assertions true in the model of $\mathrm{TT}_3$).

We further need to discuss the relationship between the results of our paper and the entirely independent work of Hazen in [5], of which we became aware after we had already obtained the results on functions described here. Hazen argues that there cannot be a general representation of binary relations in $\mathrm{TT}_3$ (which he calls "monadic third-order logic", terminology we adopted in our title) for reasons essentially similar to reasons given in our analysis. He certainly gives an accurate general description of the reasons for this fact, using the same approach of partitioning sort 0 into three-element sets and considering a function with these sets as its orbits. We are not sure that his argument is completely rigorous (it may actually be, but the style is unfamiliar to us); Hazen himself says (personal communication, quoted with his permission) that his argument looks like a Fraenkel-Mostowski construction argument for his result framed by someone who had never heard of Fraenkel-Mostowski constructions. We note that Hazen also has shown in prior work ([2]) that existence of a linear order on sort 0 is sufficient to yield a representation of binary relations in monadic third order logic.[8] We believe that we should in justice grant that Hazen has given a very similar argument for non-representability of binary relations in general prior to ours; we have made the further contributions however,

[8] And his representation of relations using a linear order does not require ten distinct sort 0 objects, but we do not regard this as a strong assumption: ours is more economical to state.

of a more rigorous presentation of a similar argument using FM model techniques, positive results concerning representation of large classes of functions and total antisymmetric relations in monadic third order logic, and proofs of non-representability in the specific cases of functions and total antisymmetric relations. Our detailed analysis of the case of total antisymmetric relations is new. Hazen has pointed out to us the relevance of our results to evaluation of the capabilities of David Lewis's logical framework exhibited in *Parts of Classes* ([8]).

We wish to acknowledge very useful conversations with Allen Hazen in the course of this work.

**References**


1. Boffa, M and Crabbé, M., "Les théorèmes 3-stratifiée de $NF_3$", C. R. Acad. Sci. Paris, t. 180 (23 June 1975).
2. Burgess, J. P., Hazen, A. P. and Lewis, D, "Appendix on pairing", in Lewis, David, *Parts of Classes*, Basil Blackwell, Oxford, 1991, pp. 121-149.
3. Fourny, Laurent, *Le nombre naturel de Frege dans les théories typées et difficultés associées*, Master's thesis, Catholic University of Louvain-la-Neuve, 2005
4. Grishin, V. N., "Consistency of a fragment of Quine's NF system", Sov. Math. Dokl. 10 (1969), pp. 1387-90.
5. Hazen, A. P. , "Relations in monadic third order logic", *Journal of Philosophical Logic*, Vol. 26, No. 6 (Dec., 1997), pp. 619-628
6. Jech, Thomas, *Set theory*, Academid Press 1978, pp. 199-201.
7. Kuratowski, C., "Sur l'operation A de l'analysis situs", *Fundam. Math.* 3 (1922), 182-199.
8. Lewis, David, *Parts of Classes*, Basil Blackwell, Oxford, 1991.
9. Oswald, Marcel, *Axiomatique de Peano et Definitions Fregeene des Nombres Naturels dans les Familles d'Ensembles*, Louvain-la-Neuve, Master's thesis, Catholic University of Louvain-la-Neuve, 1985.
10. Pabion, J.F., "$TT_3I$ est équivalent à l'arithmétique du second ordre." Comptes Rendus hebdomadaires des sances de l'Acadmie des Sciences de Paris (srie A) 290 (1980), pp. 1117-1118.
11. Quine, W. v. O., "New Foundations for Mathematical Logic", *American Mathematical Monthly*, 44 (1937), pp. 70-80.
12. Ramsey, Frank P., "The foundations of mathematics", *Proceedings of the London Mathematical Society*, vol. 25 (1925), pp. 338-384.
13. Specker, E.P. [1962] "Typical ambiguity". *Logic, methodology and philosophy of science*, ed. E. Nagel, Stanford University Press, pp. 116-123.
14. Whitehead, A.N., and Russell, B., *Principia Mathematica*, 3 vols., Cambridge University Press, 1910, 1912, and 1913.
15. Wang, H., *Logic, Computers, and Sets*, Chelsea, 1970, p. 406.
16. Wiener, N., "A simplification of the logic of relations", *Proceedings of the Cambridge Philosophical Society*, 17 (1914), pp. 387-90.

17. Zermelo, E., Neuer Beweis für die Möglichkeit einer Wohlordnung, *Mathematische Annalen* 65 (1908): 107–128.