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

M. Randall Holmes

version with further revisions: 5 pm Boise time 7/9/2017

1 Higher order logics TST and TST_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 ω, which we call TST (*theórie simple des types*, a term traditionally used by the Belgian school of students of Quine's NF for this theory). This theory is often confused with the type theory of Russell and Whitehead's [12], but is far simpler: before TST 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 [14]) and the ramifications of the type theory of [12], motivated by predicativist scruples, had to be stripped out, as by Ramsey ([10]). The history of this theory is outlined in [13]: it seems to actually first appear in print about 1930, long after [12]. We are specifically concerned with an initial segment TST_3 of this theory.

TST 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 TST are extensionality:

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

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

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

for each formula ϕ in which A does not occur free, and for each assignment of types to variables which makes sense. The witness to the instance of comprehension associated with a formula ϕ, 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 TST_n is the subtheory of TST using only the n sorts indexed by m with $0 \leq m < n$. TST_n is a formalization of nth 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 TST. The existence of the standard Kuratowski pair (for which the index reference is [6]) shows that TST_4 contains a full implementation of second order logic of binary relations: 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. More generally, TST_{3n-2} contains a full implementation of the nth order logic of binary relations [in which each "unary relation" (i.e., property) $P(x)$ could be represented by the binary relation $x\, P^*\, y$ defined as holding iff $x = y \wedge P(x)$], and TST itself is just as good an implementation of higher order logic of binary relations as of higher order logic of monadic predicates. Standard reductions of higher arity relations to binary relations via pairing establish that TST implements higher order logic of any order over relations of any arity.

We are not really concerned with relations of higher arity here, but we note that there is a specific small n such that models of TST_n with infinitely many individuals (in the sense of the metatheory: they need not satisfy an axiom of infinity) will contain a full implementation of the second order theory of m-ary relations for each concrete m. Because there are at least $m-1$ individuals, the concrete Frege natural numbers $0, \ldots, m-1$ exist in sort 2, so m-tuples can be represented as functions with domain $\{0, \ldots, m-1\}$ in sort 5, and arbitrary m-ary relations on sort 0 objects are representable in sort 6. This shows that $n = 6$ works: we are not certain that 6 is the minimal value for which this works, but we are not concerned to address this question here.

It is useful to note that there is an internal notion of *finite set* in TST_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 TST_3, where the ordered pair of Kuratowski is not available.

2 Representation of binary relations in TST_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 TST_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 TST_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 representation of well-orderings as sets in his 1908 proof of the Well-Ordering Theorem ([15]): 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 TST_3 as sort 2 sets of unordered pairs.

If there is a linear order on sort 0 in a model of TST_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 TST_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 $C(x, y)$ (which we might read "x is chosen over y"; $C(x, x)$ is always true, and if x and y are distinct, exactly one of $C(x, y)$ and $C(y, x)$ is true) 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 TST or even TST_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 TST_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 $x \neq C_v$. In this case we call (u, v) a "bad pair with respect to x" or just a bad pair. Note that bad pairs (and bad triples introduced below) are objects in the metatheory: TST_3 does not support an ordered pair.

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 $(B \setminus \{v\}) \cup \{x\} = A$.

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) or (s, x) 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) or (s, x) 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. In this situation, which occurs whichever of u, v we take s to be, we say that (x, u, v) is a "bad triple" for C.

Thus we can assert the existence of a particular kind of total antisymmetric relation (one which has no bad triples) in the language of TST_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.

3 Representation of a large class of functions in TST_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: We fix a sort 0 variable x and a sort 0 variable y. We call a formula ϕ *functional* iff $(\forall x : (\exists y : \phi) \wedge (\forall xyz : \phi \wedge \phi[z/y] \rightarrow y = z)$ holds. When ϕ 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 ϕ, 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: If ϕ is a functional formula and A is a sort 1 set, we say that A is closed under f_ϕ 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_ϕ, as the intersection of all sets which are closed under f_ϕ and contain x as an element. We define a finite cycle in f_ϕ 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: We work in an arbitrary model of TST_3. There is a uniform way to represent functional formulas ϕ by sets $[f_\phi]$ for each ϕ for which there is a choice set C_ϕ for finite cycles in f_ϕ of cardinality greater than 2.

Proof: The set $[f_\phi]$ which we take as representing the function f_ϕ 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_ϕ. It is also important to note that every forward orbit in f_ϕ is also a forward orbit of this restriction.
2. singletons of elements of C_ϕ
3. singletons of elements of f_ϕ“$C_\phi = \{y : (\exists x \in C_\phi : \phi(x, y)))\}$.

Given a set F, we indicate how to reverse engineer a functional formula ϕ 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_ϕ 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_ϕ 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_ϕ, 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_ϕ 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_ϕ"C_ϕ as a subset.

This motivates the following

Definition: 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_ϕ 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: 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_ϕ, $f_\phi(x)$ is the element of f_ϕ"C_ϕ contained in the same finite cycle in f_ϕ. This motivates the following

Definition: 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 $O(x) = \mathtt{Orbit}_F(x)$.[1]

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.

[1] 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 $> \omega$).

4 Applications: cardinality can be represented in TST_3 and 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 TST_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_ϕ 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_ϕ (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 NF_3 of Quine's New Foundations (the set theory described in [9], usually abbreviated NF) shown to be consistent by Grishin ([4]). This was shown by somewhat different methods in unpublished work by Henrard (discussed in [7], [3]). That cardinality is definable in 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 TST_3) having the same cardinality, we do have the ability to define the cardinal $|A|$ as the (sort 2 in TST_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 wiht equality and membership whose axioms are the axioms of TST with distinctions of sort between variables dropped (without creating identifications between variables); NF_n has the same relationship to TST_n. NF_4 was shown in [4] to be the same theory as NF. Any two models of TST_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 TST_3 which is externally infinite is readily shown to be elementarily equivalent to a countable model of TST_3 which is externally infinite and has the splitting property. A countable model of TST_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 TST_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 [11]. The net effect of this is that the stratified theorems of NF_3 (the ones which can be read as theorems of TST_3 by assigning sorts to variables) are in fact the theorems which hold in all externally infinite models of TST_3 (including externally infinite models of TST_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 ω. 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 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 ([8]) 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 asymmetric relations: let C be a total asymmetric 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 function which sends u to v, v to x, and x to u in each bad triple (u, v, x), using the given choice set on the bad triples. C_x can then be defined as in the represention given above, which may fail, but in the case where that definition fails, C_3 provides the needed information.

The condition asserting the existence of such a representation of a total asymmetric relation follows: there are sets D, E, and F such that for each x there is either a unique D_x satisfying the conditions stated in the earlier partial representation or there is a triple of sets $A, B, C \in D$ and u, v such that A, B, C are obtained by taking the union of the same set D_x^-, not containing any of x, u, v, with one of the two element subsets of $\{x, u, v\}$. We refer to $\{x, u, v\}$ as a bad triple in this case. For any distinct x, y, exactly one of $x \in D_y$ and $y \in D_x$ holds: 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 $D_x^- \neq D_y^-$. F is the set representation of a bijection whose domain is the union of the bad triples, which has no fixed points, and which permutes each bad triple. The relation $C(x, y)$ is defined as "$x \in D_y \vee y = F[x] \vee y = x$". Of course, if this condition holds we can define a complete implementation of binary relations.

Note that under the hypothesis 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 TST_3 the assertion that there is a total antisymmetric relation. We are not saying that 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 TST_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 TST_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: We say that a formal implementation of functions in TST_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 TST_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 TST_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 TST_3 with the addition of axioms that $F(x, y)$ is a functional formula and that all instances of the comprehension scheme for TST_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.[2]

[2]The single variable f may be replaced throughout by a finite vector $f_1, \ldots, f_n$, if the

Definition: Similarly, we say that a formal implementation of total antisymmetric relations in TST_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 TST_3 with 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 $\mathtt{tarel}_R$: we will usually write it $\mathtt{tarel}_R(r)$ in order to signal this. $\mathtt{tarelapp}$ is a formula in the language of TST_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 $\mathtt{tarelapp}(r, x, y)$ to signal this. In the extension of TST_3 with the addition of axioms that $R(x, y)$ is a total antisymmetric relation and that all instances of the comprehension scheme for TST_3 involving the new primitive relation R hold, with no other additional axioms, we require that $(\exists r : \mathtt{tarel}_R(r))$ is a theorem and that $\mathtt{tarel}_R(r) \rightarrow (\mathtt{tarelapp}(r, x, y) \leftrightarrow R(x, y))$ is a theorem.[3]

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 TST_3. The intended sense of $\mathtt{fun}_F(f)$ is "f is the set implementation of the functional binary relation F"; the intended sense of $\mathtt{app}(f, x, y)$ is "y is the result of applying the function represented by the set f to x". The intended sense of $\mathtt{tarel}_R(r)$ is "r is the set implementation of the total antisymmetric relation R"; the intended sense of $\mathtt{tarelapp}(r, x, y)$ 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. At this point we stipulate that 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.

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.

[3]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.

We also note that any model of TST_3 in which the set implementing sort 0 is not larger than the collection of atoms is isomorphic to a model of TST_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 TST_3 a "natural model" of TST_3 in ZFA.

Theorem: There is no formal implementation of functions in TST_3, nor is there any formal representation of total antisymmetric relations in TST_3.

Proof: We set out to construct a natural model of TST_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 π 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 TST_4 will consist of the set of atoms already mentioned. Sort 1 of our model of TST_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 TST_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 TST_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 TST_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 TST_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 $|B \cap o| = g(o)$.

Now observe (it is evident from the descriptions of sort 1 and sort 2 sets) that the model of TST_3 consisting of sorts 0,1,2 of the model of TST_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 TST_4 defined with the smaller group G.

By examination of the model of TST_3 just described as an initial segment of the model of TST_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 π, which is impossible.[4]

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 TST_4 and in the model of TST_3 which is its initial segment. No formulas $\mathtt{tarel}_R(r)$ and $\mathtt{tarelapp}(R, x, y)$ in the language of TST_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 that given above. Let r satisfy $\mathtt{tarel}_R(r)$ where R denotes the total asymmetric 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 π, which is impossible.[5]

This has a corollary with an ironic flavor: if we provide a predicate R

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

[5]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 $r_1, \ldots, r_n$.

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 in the language of TST_3: we have shown above that we can express the assertion that there is some total antisymmetric relation if we have the additional hypothesis AC_3 that each collection of disjoint three-element sets has a choice set.

We make a final remark about choice principles in this model. The choice principle AC_2 = "every disjoint collection of pairs has a choice set" holds in the model of TST_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 TST_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 TST_3, 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 theses [7], [3]; we became aware of it because Henrard's results are folklore among the small community of NF researchers. 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 TST_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 TST_3 is elementarily equivalent to a model of NF_3 (in the sense that the stratified assertions true in the model of NF_3 correspond exactly to the assertions true in the model of TST_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 described here. Hazen argues that there cannot be a general representation of binary relations in TST_3 (which he calls "monadic third-order logic") 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) 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. 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, of a more rigorous presentation of a similar argument using FM model techniques, positive results concerning representation of large classes of functions and total asymmetric relations in monadic third order logic, and proofs of non-representability in the specific cases of functions and total asymmetric relations.

References

[1] Boffa, M and Crabbé, M., "Les théorèmes 3-stratifié 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] Kuratowski, C., "Sur l'operation A de l'analysis situs", *Fundam. Math.* 3 (1922), 182-199.

[7] 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.

[8] 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.

[9] Quine, W. v. O., "New Foundations for Mathematical Logic", *American Mathematical Monthly*, 44 (1937), pp. 70-80.

[10] Ramsey, Frank P., "The foundations of mathematics", *Proceedings of the London Mathematical Society*, vol. 25 (1925), pp. 338-384.

[11] Specker, E.P. [1962] "Typical ambiguity". *Logic, methodology and philosophy of science*, ed. E. Nagel, Stanford University Press, pp. 116-123.

[12] Whitehead, A.N., and Russell, B., *Principia Mathematica*, 3 vols., Cambridge University Press, 1910, 1912, and 1913.

[13] Wang, H., *Logic, Computers, and Sets*, Chelsea, 1970, p. 406.

[14] Wiener, N., "A simplification of the logic of relations", *Proceedings of the Cambridge Philosophical Society*, 17 (1914), pp. 387-90.

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