New Foundations is consistent

M. Randall Holmes

November 14, 2023

Contents

1 Remarks on this version

This document is probably my best overall version so far. The immediate occasion for its preparation was to serve students attempting to verify the proof in Lean. A formal verification should avoid metamathematics, so it is the fact that the structure defined in section 3 is a model of TTT which should be verified, and further, a finite axiomatization (mod type indexing) of TST and thus TTT should be verified in the model in lieu of the usual statement of the axiom of comprehension of TTT.

Versions after 8/11/2022 use a different set theoretical coding of structures for models of TTT, but the argument remains about the same. The last version with the older coding remains up on my web site for comparison.

1.1 Version updates

5/18/2023: We thank a careful reader for noticing that some definitions from the previous version had not been carried forward. The required definitions have been added (of the notion of a symmetric element of τ^*_α, of the notation (α, β, B) for the unique element x of τ^*_α which is extensional and has $x \cap \tau_\beta = B$, and of the notation $S_{(\beta)}$ for a certain operation on supports, which now appears in the reprise of strong supports in the appropriate place).

3/31/2023: Some editing motivated by Sky Wilshaw's concern about bookkeeping difficulties she is encountering the proof of Freedom of Action having to do with enforcing the condition that the process of extension of approximations in the proof preserves injectivity. The difficulty she is encountering seems difficult to describe and I may improve the language about it here. I made various technical corrections through the entire text of the construction through the end of Freedom of Action, fixing various notational bugs. There is some language about clarifying the inductive hypothesis in the way she needs, which doesn't satisfy me perfectly.

A notational point in Freedom of Action is that I'm systematically using the notation $\pi^*_A(x)$ for values of $\pi_A(x)$ computed prior to the complete determination of π_A and π.

2/10/2023: A provisional clarification of the proof of Freedom of Action (that the same $\pi_{L,M}$'s are used in calculation of lower index allowable permutations) is given: this should be revisited. This issue came up in the formal proof.

A further change: the maps $\pi_{L,M}$ are given an exact definition, which ensures that sensible coherence properties will obtain. The way I was doing it before this was an error: it didnt impose enough restrictions to ensure that recursive calls to the freedom of action construction worked as desired. It's an error local to this version: previous approaches defined the $\pi_{L,M}$'s in a more concrete and limited way. On p. 19 I define the well-ordering $<_L$ on any litter L (ordering by the last index). I then use this to define the explicit $\pi_{L,M}$'s in the freedom of action proof, p. 29.

1/27/2023: Expanded the description of the back-and-forth construction of the order built on each type just after it is constructed. Also, Sky points out that I need to say that near-litters in designated supports are litters.

1/20/2023: Very small edits.

1/13/2022: Continuing work on the Lean project. Identified some typos to be fixed.

10/30/2022: Minor edits, intending to post to arxiv

5/18/2022 (continuing 5/19): Proofreading and intending to strip out comments addressed to the Lean group (so it looks more like a proper paper draft again).

9/13/2022: Proofreading, in some cases quite important (the definition of β-approximation was missing an important condition assumed in the proof). A second update with rephrasing of the definition of approximation using the notion of condition.

8/14/2022: Cleaning this up for posting on Arxiv to replace a version of the previous family. We note for readers on Arxiv that this version does two things: it corrects an actual omission of an important case in the proofs in the last version of the previous family, and at the same time makes a systematic change in the set theoretic implementation of structures for the language of TTT that we use. The corrections needed for the error of omission in the previous version of the proof are relatively minor; the changes required for the use of different data structures to represent structures for the language of TTT are considerable but do not really affect the way the argument works. I hope that they will improve intelligibility.

This version still contains extensive notes for the Lean project workers pointing out where the corrections for the error of omission appear; these, and most or all comments addressed to the Lean project workers, will probably be taken out in the next update.

8/11/2022: Fixed the last section on the axiom of union: the $\gamma = -1$ issue requires a somewhat more careful (though basically similar) approach.

8/10/2022: Posting typo fixes. There are a lot of them, this will probably happen several times (9:34 am Boise time).

8/9/2022, alpha release: Everything is processed into the new framework. I believe the whole argument works, though I think some rephrasing is needed in the last part. I'm posting this as a prospective replacement of the web paper, but leaving the old version for now, as I am certain that reconciliation issues will arise and the old version will be needed for comparison.

8/8/2022, pre-release: Translating into a different framework and correcting an omission in the definition of dependencies of conditions. At this point a large block of material in section 4 remains to be translated into the new framework and checked for problems with the $\gamma = -1$ fix.

9/26/2023: Revised the definition of ordered strong support to make it clear that ordered strong supports do not have to have their native order agree in all cases with the global order on support conditions, though the two orders will agree when a dependency condition exists. Further fine revisions later on 9/26 clarifying the same point.

9/27/2023: Made a correction to the definition of specification: if a specification is to be a well-ordering of specification conditions, they all have to be different. Adding position as a component of atom and flexible litter specification conditions does the triclk. It would also work to let a specification be a function from an ordinal less than κ to specification conditions as originally defined.

10/5/2023: Technical improvements to the argument on counting model elements due to the diligence of Sky Wilshaw, both in generating Lean code and in asking questions persistently and holding my feet to the fire.

10/10/2023: Corrected a material error in the argument on counting model elements, thanks to Sky for leaning on us about this. I think this was an error introduced in the editing of this version from the previous one, and so far uninspected because the Lean proof effort has only recently reached this part of the paper.

Later on 10/10: the first version of the fix did not work. Notice that this further refinement involves imposing a condition on choice of designated supports which has the effect that if x and y are in the same orbit in a type under allowable permutations, their designated supports will be as well (already proposed by Sky in the Lean development).

11/7/2023: an experimental version which makes a major revision, still propagating through the text. The f maps take as input not simply elements of types but elements of types with supports. This essentially removes the need for strong supports.

11/14/2023: Changes propagated into the central part of the counting argument. It seems that the notion of strong or even nice support is no longer needed.

11/14/2023 (after Zoom meeting): Further edits and indication of a gap to be filled in the theorem relatiing specifications and orbits of supports, following our conversation in the NF Zoom meeting.

2 Development of relevant theories

2.1 The simple theory of types TST and TSTU

We introduce a theory which we call the simple typed theory of sets or TST, a name favored by the school of Belgian logicians who studied NF (*théorie simple des types*). This is not the same as the simple type theory of Ramsey and it is most certainly not Russell's type theory (see historical remarks below).

TST is a first order multi-sorted theory with sorts (types) indexed by the nonnegative integers. The primitive predicates of TST are equality and membership.

The type of a variable x is written $\texttt{type}(`x')$: this will be a nonnegative integer. A countably infinite supply of variables of each type is supposed. An atomic equality sentence '$x = y$' is well-formed iff $\texttt{type}(`x') = \texttt{type}(`y')$. An atomic membership sentence '$x \in y$' is well-formed iff $\texttt{type}`(x') + 1 = \texttt{type}(`y')$.

The axioms of TST are extensionality axioms and comprehension axioms.

The extensionality axioms are all the well-formed assertions of the shape $(\forall xy : x = y \leftrightarrow (\forall z : z \in x \leftrightarrow z \in y))$. For this to be well typed, the variables x and y must be of the same type, one type higher than the type of z.

The comprehension axioms are all the well-formed assertions of the shape $(\exists A : (\forall x : x \in A \leftrightarrow \phi))$, where ϕ is any formula in which A does not occur free.

The witness to $(\exists A : (\forall x : x \in A \leftrightarrow \phi))$ is unique by extensionality, and we introduce the notation $\{x : \phi\}$ for this object. Of course, $\{x : \phi\}$ is to be assigned type one higher than that of x; in general, term constructions will have types as variables do.

The modification which gives TSTU (the simple type theory of sets with urelements) replaces the extensionality axioms with the formulas of the shape

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

allowing many objects with no elements (called atoms or urelements) in each positive type. A technically useful refinement adds a constant $\emptyset^i$ of each positive type i with no elements: we can then address the problem that $\{x^i : \phi\}$ is not uniquely defined when ϕ is uniformly false by defining $\{x^i : \phi\}$ as $\emptyset^{i+1}$ in this case.

2.1.1 Typical ambiguity

TST(U) exhibits a symmetry which is important in the sequel.

Provide a bijection $(x \mapsto x^+)$ from variables to variables of positive type satisfying $\texttt{type}(x^+) = \texttt{type}(x) + 1$.

If ϕ is a formula, define ϕ^+ as the result of replacing every variable x (free and bound) in ϕ with x^+ (and occurrences of $\emptyset^i$ with $\emptyset^{i+1}$ if this is in use). It should be evident that if ϕ is well-formed, so is ϕ^+, and that if ϕ is a theorem, so is ϕ^+ (the converse is not the case). Further, if we define a mathematical object as a set abstract $\{x : \phi\}$ we have an analogous object $\{x^+ : \phi^+\}$ of the next higher type (this process can be iterated).

The axiom scheme asserting $\phi \leftrightarrow \phi^+$ for each closed formula ϕ is called the Ambiguity Scheme. Notice that this is a stronger assertion than is warranted by the symmetry of proofs described above.

2.1.2 Historical remarks

TST is not the type theory of the *Principia Mathematica* of Russell and Whitehead ([15]), though a description of TST is a common careless description of Russell's theory of types.

Russell described something like TST informally in his 1904 *Principles of Mathematics* ([14]). The obstruction to giving such an account in *Principia Mathematica* was that Russell and Whitehead did not know how to describe ordered pairs as sets. As a result, the system of *Principia Mathematica* has an elaborate system of complex types inhabited by n-ary relations with arguments of specified previously defined types, further complicated by predicativity restrictions (which are in effect cancelled by an axiom of reducibility). The simple theory of types of Ramsey eliminates the predicativity restrictions and the axiom of reducibility, but is still a theory with complex types inhabited by n-ary relations.

Russell noticed a phenomenon like the typical ambiguity of TST in the more complex system of *Principia Mathematica*, which he refers to as "systematic ambiguity".

In 1914 ([22]), Norbert Wiener gave a definition of the ordered pair as a set (not the one now in use) and seems to have recognized that the type theory of *Principia Mathematica* could be simplified to something like TST, but he did not give a formal description. The theory we call TST was apparently first described by Tarski in 1930 ([20]).

It is worth observing that the axioms of TST look exactly like those of "naive set theory", the restriction preventing paradox being embodied in the restriction of the language by the type system. For example, the Russell paradox is averted because one cannot have $\{x : x \notin x\}$ because $x \in x$ (and so its negation $\neg x \in x$) cannot be a well-formed formula.

It was shown around 1950 (in [8]) that Zermelo set theory proves the consistency of TST with the axiom of infinity; TST + Infinity has the same consistency strength as Zermelo set theory with separation restricted to bounded formulas.

2.2 Some mathematics in TST; the theories TST_n and their natural models

We briefly discuss some mathematics in TST.

We indicate how to define the natural numbers. We use the definition of Frege (n is the set of all sets with n elements). 0 is $\{\emptyset\}$ (notice that we get a natural number 0 in each type $i+2$; we will be deliberately ambiguous in this discussion, but we are aware that anything we define is actually not unique, but reduplicated in each type above the lowest one in which it can be defined). For any set A at all we define $\sigma(A)$ as $\{a \cup \{x\} : a \in A \wedge x \notin a\}$. This is definable for any A of type $i+2$ (a being of type $i+1$ and x of type i). Define 1 as $\sigma(0)$, 2 as $\sigma(1)$, 3 as $\sigma(2)$, and so forth. Clearly we have successfully defined 3 as the set of all sets with three elements, without circularity. But further, we can define $\mathbb{N}$ as $\{n : (\forall I : 0 \in I \wedge (\forall x \in I : \sigma(x) \in I) \rightarrow n \in I\}$, that is, as the intersection of all inductive sets. $\mathbb{N}$ is again a typically ambiguous notation: there is an object defined in this way in each type $i+3$.

The collection of all finite sets can be defined as $\bigcup \mathbb{N}$. The axiom of infinity can be stated as $V \notin \bigcup \mathbb{N}$ (where $V = \{x : x = x\}$ is the typically ambiguous symbol for the type $i+1$ set of all type i objects). It is straightforward to show that the natural numbers in each type of a model of TST with Infinity are isomorphic in a way representable in the theory.

Ordered pairs can be defined following Kuratowski and a quite standard theory of functions and relations can be developed. Cardinal and ordinal numbers can be defined as Frege or Russell would have defined them, as isomorphism classes of sets under equinumerousness and isomorphism classes of well-orderings under similarity.

The Kuratowski pair $(x, y) = \{\{x\}, \{x, y\}\}$ is of course two types higher than its projections, which must be of the same type. There is an alternative definition (due to Quine in [11]) of an ordered pair $\langle x, y\rangle$ in TST + Infinity which is of the same type as its projections x, y. This is a considerable technical convenience but we will not need to define it here. Note for example that if we use the Kuratowski pair the cartesian product $A \times B$ is two types higher than A, B, so we cannot define $|A| \cdot |B|$ as $|A \times B|$ if we want multiplication of cardinals to be a sensible operation. Let ι be the singleton operation and define $T(|A|)$ as $|\iota``A|$ (this is a very useful operation sending cardinals of a given type to cardinals in the next higher type which seem intuitively to be the same; also, it is clearly injective, so has a (partial) inverse operation T^{-1}). The definition of cardinal multiplication if we use the Kuratowski pair is then $|A| \cdot |B| = T^{-2}(|A \times B|)$. If we use the Quine pair this becomes the usual definition $|A| \cdot |B| = |A \times B|$. Use of the Quine pair simplifies matters in this case, but it should be noted that the T operation remains quite important (for example it provides the internally representable isomorphism between the systems of natural numbers in each sufficiently high type).

Note that the form of Cantor's Theorem in TST is not $|A| < |\mathcal{P}(A)|$, which would be ill-typed, but $|\iota``A| < |\mathcal{P}(A)|$: a set has fewer unit subsets than subsets. The exponential map $\exp(|A|) = 2^{|A|}$ is not defined as $|\mathcal{P}(A)|$, which would be one type too high, but as $T^{-1}(|\mathcal{P}(A))$, the cardinality of a set $X$ such that $|\iota``X| = |\mathcal{P}(A)|$; notice that this is partial. For example $2^{|V|}$ is not defined (where $V = \{x : x = x\}$, an entire type), because there is no X with $|\iota``X| = |\mathcal{P}(V)|$, because $|\iota``V| < |\mathcal{P}(V)| \leq |V|$, and of course there is no set larger than V in its type.

For each natural number n, the theory TST_n is defined as the subtheory of TST with vocabulary restricted to use variables only of types less than n (TST with n types). In ordinary set theory TST and each theory TST_n have natural models, in which type 0 is implemented as a set X and each type i in use is implemented as $\mathcal{P}^i(X)$. It should be clear that each TST_n has natural models in bounded Zermelo set theory, and TST has natural models in a modestly stronger fragment of ZFC.

Further, each TST_n has natural models in TST itself, though some care must be exercised in defining them. Let X be a set. Implement type i for each $i < n$ as $\iota^{(n-1)-i}``\mathcal{P}^i(X)$. If X is in type j, each of the types of this interpretation of TST_n is a set in the same type $j + n - 1$. For any relation R, define R^ι as $\{(\{x\}, \{y\}) : xRy\}$. The membership relation of type $i - 1$ in type i in the interpretation described is the restriction of $\subseteq^{\iota^{(n-1)-i}}$ to the product of the sets implementing type $i - 1$ and type i.

Notice then that, for each concrete natural number n, we can define truth for formulas in these natural models of TST_n in TST, though not in a uniform way which would allow us to define truth for formulas in TST in TST.

Further, both in ordinary set theory and in TST, observe that truth of sentences in natural models of TST_n is completely determined by the cardinality of the set used as type 0. since two natural models of TST or TST_n with base types implemented by sets of the same cardinality are clearly isomorphic.

2.3 New Foundations and NFU

In [12], 1937, Willard van Orman Quine proposed a set theory motivated by the typical ambiguity of TST described above. The paper in which he did this was titled “New foundations for mathematical logic”, and the set theory it introduces is called “New Foundations” or NF, after the title of the paper.

Quine’s observation is that since any theorem ϕ of TST is accompanied by theorems $\phi^+, \phi^{++}, \phi^{+++}, \ldots$ and every defined object $\{x : \phi\}$ is accompanied by $\{x^+ : \phi^+\}, \{x^{++} : \phi^{++}\}, \{x^{+++} : \phi^{+++}\}$, so the picture of what we can prove and construct in TST looks rather like a hall of mirrors, we might reasonably (?) suppose that the types are all the same.

The concrete implementation follows. NF is the first order unsorted theory with equality and membership as primitive with an axiom of extensionality $(\forall xy : x = y \leftrightarrow (\forall z : z \in x \leftrightarrow z \in y))$ and an axiom of comprehension $(\exists A : (\forall x : x \in A \leftrightarrow \phi))$ for each formula ϕ in which A is not free which can be obtained from a formula of TST by dropping all distinctions of type. We give a precise formalization of this idea: provide a bijective map $(x \mapsto x^*)$ from the countable supply of variables (of all types) of TST onto the countable supply of variables of the language of NF. Where ϕ is a formula of the language of TST, let ϕ^* be the formula obtained by replacing every veriable x, free and bound, in ϕ with x^*. For each formula ϕ of the language of TST in which A is not free in ϕ^* and each variable x^*, an axiom of comprehension of NF asserts $(\exists A : (\forall x^* : x^* \in A \leftrightarrow \phi^*))$.

In the original paper, this is expressed in a way which avoids explicit dependence on the language of another theory. Let ϕ be a formula of the language of NF. A function σ is a stratification of ϕ if it is a (possibly partial) map from variables to non-negative integers such that for each atomic subformula ‘$x = y$’ of ϕ we have $\sigma(‘x’) = \sigma(‘y’)$ and for each atomic subformula ‘$x \in y$’ of ϕ we have $\sigma(‘x’)+1 = \sigma(‘y’)$. A formula ϕ is said to be stratified iff there is a stratification of ϕ. Then for each stratified formula ϕ of the language of NF and variable x we have an axiom $(\exists A : (\forall x : x \in A \leftrightarrow \phi))$. The stratified formulas are exactly the formulas ϕ^* up to renaming of variables.

NF has been dismissed as a “syntactical trick” because of the way it is defined. It might go some way toward dispelling this impression to note that the stratified comprehension scheme is equivalent to a finite collection of its instances, so the theory can be presented in a way which makes no reference to types at all. This is a result of Hailperin ([3]), refined by others. One obtains a finite axiomatization of NF by analogy with the method of finitely axiomatizing von Neumann-Gödel-Bernays predicate class theory. It should further be noted that the first thing one does with the finite axiomatization is prove stratified comprehension as a meta-theorem, in practice, but it remains significant that the theory can be axiomatized with no reference to types at all.

For each stratified formula ϕ, there is a unique witness to

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

(uniqueness follows by extensionality) whch we denote by $\{x : \phi\}$.

Jensen in [10], 1969 proposed the theory NFU which replaces the extensionality axiom of NF with

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

allowing many atoms or urelements. One can reasonably add an elementless constant $\emptyset$, and define $\{x : \phi\}$ as $\emptyset$ when ϕ is false for all x.

Jensen showed that NFU is consistent and moreoever NFU + Infinity + Choice is consistent. We will give an argument similar in spirit though not the same in detail for the consistency of NFU in the next section.

An important theorem of Specker ([19], 1962) is that NF is consistent if and only if TST + the Ambiguity Scheme is consistent. His method of proof adapts to show that NFU is consistent if and only if TSTU + the Ambiguity Scheme is consistent. Jensen used this fact in his proof of the consistency of NFU. We indicate a proof of Specker's result using concepts from this paper below.

In [18], 1954, Specker had shown that NF disproves Choice, and so proves Infinity. At this point if not before it was clear that there is a serious issue of showing that NF is consistent relative to some set theory in which we have confidence. There is no evidence that NF is any stronger than TST + Infinity, the lower bound established by Specker's result.

Note that NF or NFU supports the implementation of mathematics in the same style as TST, but with the representations of mathematical concepts losing their ambiguous character. The number 3 really is realized as the unique set of all sets with three elements, for example. The universe is a set and sets make up a Boolean algebra. Cardinal and ordinal numbers can be defined in the manner of Russell and Whitehead.

The apparent vulnerability to the paradox of Cantor is an illusion. Applying Cantor's theorem to the cardinality of the universe in NFU gives $|\iota``V| < |\mathcal{P}(V)| \leq |V|$ (the last inequality would be an equation in NF), from which we conclude that there are fewer singletons of objects than objects in the universe. The operation $(x \mapsto \{x\})$ is not a set function, and there is every reason to expect it not to be, as its definition is unstratified. The resolution of the Burali-Forti paradox is also weird and wonderful in NF(U), but would take us too far afield.

2.4 Tangled type theory TTT and TTTU

In [5], 1995, this author described a reduction of the NF consistency problem to consistency of a typed theory, motivated by reverse engineering from Jensen's method of proving the consistency of NFU.

Let λ be a limit ordinal. It can be ω but it does not have to be.

In the theory TTT (tangled type theory) which we develop, each variable x is supplied with a type $\texttt{type}(`x') < \lambda$; we are provided with countably many distinct variables of each type.

For any formula ϕ of the language of TST and any strictly increasing sequence s in λ, let ϕ^s be the formula obtained by replacing each variable of type i with a variable of type $s(i)$. To make this work rigorously, we suppose that we have a bijection from type i variables of the language of TST to type α variables of the language of TTT for each natural number i and ordinal $\alpha < \lambda$.

TTT is then the first order theory with types indexed by the ordinals below λ whose well formed atomic sentences '$x = y$' have $\texttt{type}(`x') = \texttt{type}(`y')$ and whose atomic sentences '$x \in y$' satisfy $\texttt{type}(`x') < \texttt{type}(`y')$, and whose axioms are the sentences ϕ^s for each axiom ϕ of TST and each strictly increasing sequence s in λ. TTTU has the same relation to TSTU (with the addition of constants $\emptyset^{\alpha,\beta}$ for each $\alpha < \beta < \lambda$ such that $(\forall \mathbf{x}_0^\alpha : \mathbf{x}_0^\alpha \not\in \emptyset^{\alpha,\beta})$ is an axiom).

It is important to notice how weird a theory TTT is. This is not cumulative type theory. Each type β is being interpreted as a power set of *each* lower type α. Cantor's theorem in the metatheory makes it clear that most of these power set interpretations cannot be honest.

There is now a striking

Theorem (Holmes): TTT(U) is consistent iff NF(U) is consistent.

Proof: Suppose NF(U) is consistent. Let (M, E) be a model of NF(U) (a set M with a membership relation E). Implement type α as $M \times \{\alpha\}$ for each $\alpha < \lambda$. Define $E_{\alpha,\beta}$ for $\alpha < \beta$ as $\{((x, \alpha), (y, \beta)) : xEy\}$. This gives a model of TTT(U). Empty sets in TTTU present no essential additional difficulties.

Suppose TTT(U) is consistent, and so we can assume we are working with a fixed model of TTT(U). Let Σ be a finite set of sentences in the language of TST(U). Let n be the smallest type such that no type n variable occurs in any sentence in Σ. We define a partition of the n-element subsets of λ. Each $A \in [\lambda]^n$ is put in a compartment determined by the truth values of the sentences ϕ^s in our model of TTT(U), where $\phi \in \Sigma$ and $\texttt{rng}(s\lceil\{0, \ldots, n-1\}) = A$. By Ramsey's theorem, there is a homogeneous set $H \subseteq \lambda$ for this partition, which includes the range of a strictly increasing sequence h. There is a complete extension of TST(U) which includes ϕ iff the theory of our model of TTT(U) includes ϕ^h. This extension satisfies $\phi \leftrightarrow \phi^+$ for each $\phi \in \Sigma$. But this implies by compactness that the full Ambiguity Scheme $\phi \leftrightarrow \phi^+$ is consistent with TST(U), and so that NF(U) is consistent by the 1962 result of Specker.

We note that we can give a treatment of the result of Specker (rather different from Specker's own) using TTT(U). Note that it is easy to see that if we have a model of TST(U) augmented with a Hilbert symbol (a primitive term construction $(\epsilon x : \phi)$ (same type as x) with axiom scheme $\phi[(\epsilon x : \phi)/x] \leftrightarrow (\exists x : \phi)$) which cannot appear in instances of comprehension (the quantifiers are not defined in terms of the Hilbert symbol, because they do need to appear in instances of comprehension) and Ambiguity (for all formulas, including those which mention the Hilbert symbol) then we can readily get a model of NF, by constructing a term model using the Hilbert symbol in the natural way, then identifying all terms with their type-raised versions. All statements in the resulting type-free theory can be decided by raising types far enough (the truth value of an atomic sentence $(\epsilon x : \phi)\, R\, (\epsilon y : \psi)$ in the model of NF is determined by raising the type of both sides (possibly by different amounts) until the formula is well-typed in TST and reading the truth value of the type raised version; R is either $=$ or $\in$). Now observe that a model of TTT(U) can readily be equipped with a Hilbert symbol if this creates no obligation to add instances of comprehension containing the Hilbert symbol (use a well-ordering of the set implementing each type to interpret a Hilbert symbol $(\epsilon x : \phi)$ in that type as the first x such that ϕ), and the argument above for consistency of TST(U) plus Ambiguity with the Hilbert symbol goes through.

Theorem (essentially due to Jensen): NFU is consistent.

Proof: It is enough to exhibit a model of TTTU. Suppose $\lambda > \omega$. Represent type α as $V_{\omega+\alpha} \times \{\alpha\}$ for each $\alpha < \lambda$ ($V_{\omega+\alpha}$ being a rank of the usual cumulative hierarchy). Define $\in_{\alpha,\beta}$ for $\alpha < \beta < \lambda$ as

$$\{((x, \alpha), (y, \beta)) : x \in V_{\omega+\alpha} \wedge y \in V_{\omega+\alpha+1} \wedge x \in y\}.$$

This gives a model of TTTU in which the membership of type α in type β interprets each (y, β) with $y \in V_{\omega+\beta} \setminus V_{\omega+\alpha+1}$ as an urelement.

Our use of $V_{\omega+\alpha}$ enforces Infinity in the resulting models of NFU (note that we did not have to do this: if we set $\lambda = \omega$ and interpret type α using V_α we prove the consistency of NFU with the negation of Infinity). It should be clear that Choice holds in the models of NFU eventually obtained if it holds in the ambient set theory.

This shows in fact that mathematics in NFU is quite ordinary (with respect to stratified sentences), because mathematics in the models of TSTU embedded in the indicated model of TTTU is quite ordinary. The notorious ways in which NF evades the paradoxes of Russell, Cantor and Burali-Forti can be examined in actual models and we can see that they work and how they work (since they work in NFU in the same way they work in NF).

Of course Jensen did not phrase his argument in terms of tangled type theory. Our contribution here was to reverse engineer from Jensen's original argument for the consistency of NFU an argument for the consistency of NF itself, which requires additional input which we did not know how to supply (a proof of the consistency of TTT itself). An intuitive way to say what is happening here is that Jensen noticed that it is possible to skip types in a certain sense in TSTU in a way which is not obviously possible in TST itself; to suppose that TTT might be consistent is to suppose that such type skipping is also possible in TST.

2.4.1 How internal type representations unfold in TTT

We have seen above that TST can internally represent TST_n. An attempt to represent types of TTT internally to TTT has stranger results. The development of the model does not depend on reading this section.

In TST the strategy for representing type i in type $n \geq i$ is to use the $n-i$-iterated singleton of any type i object x to represent x; then membership of representations of type $i-1$ objects in type i objects is represented by the relation on $n-i$-iterated singletons induced by the subset relation and with domain restricted to $n-(i+1)$-fold singletons. This is described more formally above.

In TTT the complication is that there are numerous ways to embed type α into type β for $\alpha < \beta$ along the lines just suggested. We define a generalized iterated singleton operation: where A is a finite subset of λ, ι_A is an operation defined on objects of type $\mathtt{min}(A)$. $\iota_{\{\alpha\}}(x) = x$. If A has $\alpha < \beta$ as its two smallest elements, $\iota_A(x)$ is $\iota_{A_1}(\iota_{\alpha,\beta}(x))$, where A_1 is defined as $A \setminus \{\mathtt{min}(A)\}$ (a notation we will continue to use) and $\iota_{\alpha,\beta}(x)$ is the unique type β object whose only type α element is x.

Now for any nonempty finite $A \subseteq \lambda$ with minimum α and maximum β. the range of ι_A is a set, and a representation of type α in type β. For simplicity we carry out further analysis in types $\beta, \beta+1, \beta+2\ldots$ though it could be done in more general increasing sequences. Use the notation τ_A for the range of ι_A, for each set A with β as its maximum. Each such set has a cardinal $|\tau_A|$ in type $\beta+2$. It is a straightforward argument in the version of TST with types taken from A and a small finite number of types $\beta+i$ that $2^{|\tau_A|} = |\tau_{A_1}|$ for each A with at least two elements. The relevant theorem in TST is that $2^{|\iota^{n+1}``X|} = \iota^n``X$, relabelled with suitable types from λ. We use the notation $\exp(\kappa)$ for 2^κ to support iteration. Notice that for any τ_A we have $\exp^{|A|-1}(|\tau_A|) = |\tau_{\{\beta\}}|$, the cardinality of type β. Now if A and A' have the same minimum α and maximum β but are of different sizes, we see that $|\tau_A| \neq |\tau_{A'}|$, since one has its $|A|-1$-iterated exponential equal to $|\tau_{\{\beta\}}|$ and the other has its $|A'|-1$-iterated exponential equal to $|\tau_{\{\beta\}}|$. This is odd because there is an obvious external bijection between the sets τ_A and $\tau_{A'}$: we see that this external bijection cannot be realized as a set. τ_A and $\tau_{A'}$ are representations of the same type, but this is not obvious from inside TTT. We recall that we denote $A \setminus \{\mathtt{min}(A)\}$ by A_1; we further denote $(A_i)_1$ as A_{i+1}. Now suppose that A and B both have maximum β and $A \setminus A_i = B \setminus B_i$, where $i < |A| \leq |B|$. We observe that for any concrete sentence ϕ in the language of TST_i, the truth value of ϕ in natural models with base type of sizes $|\tau_A|$ and $|\tau_B|$ will be the same, because the truth values we read off are the truth values in the model of TTT of versions of ϕ in exactly the same types of the model (truth values of ϕ^s for any s having $A \setminus A_i = B \setminus B_i$ as the range of an initial segment). This much information telling us that τ_{A_j} and τ_{B_j} for $j < i$ are representations of the same type is visible to us internally, though the external isomorphism is not. We can conclude that the full first-order theories of natural models of TST_i with base types $|\tau_A|$ and $|\tau_B|$ are the same as seen inside the model of TTT, if we assume that the natural numbers of our model of TTT are standard.

2.4.2 Tangled webs of cardinals: a suggestion of another approach not followed here

Nothing in the construction of a model of tangled type theory and verification that it is a model which appears below depends on anything in this section.

It is straightforward to transform a model of TST into a model of bounded Zermelo set theory (Mac Lane set theory) with atoms or without foundation (this depends on how type 0 is handled). Specify an interpretation of type 0 either as a set of atoms or a set of self-singletons. Then interpret type $i+1$ as inhabited by sets of type i objects in the obvious way, identifying type $i+1$ objects with objects of lower type which happen to have been assigned the same extension.

In a model of TTT, do this along some increasing sequence of types of order type ω whose range includes an infinite ordinal α. In the resulting model of bounded Zermelo set theory, let τ_A represent the cardinality of the range of ι_A as in the previous discussion (for nonempty subsets of type A all with maximum the same infinite α). Suppose further for the sake of argument that our model of TTT is λ-complete, in the sense that any subset of a type of cardinality that of λ or less is implemented as a set in each higher type. It will follow that $A \mapsto \tau_A$ is actually a function. [It is an incidentally interesting fact that the models we construct (with no dependence on this section) actually have this completeness property].

We describe the situation which holds for these cardinals.

We work in Mac Lane set theory. Choice is not assumed, and we use the Scott definition of cardinals.

Definition: If A is a nonempty finite set of ordinals which is sufficiently large, we define A_1 as $A \setminus \mathtt{min}(A)$ and A_0 as A, A_{i+1} as $(A_i)_1$.

Definition: A tangled web of cardinals of order α (an infinite ordinal) is a function τ from the set of nonempty sets of ordinals with α as maximum to cardinals such that

1. If $|A| > 1$, $\tau(A_1) = 2^{\tau(A)}$.
2. If $|A| \geq n$, the first order theory of a natural model of TST_n with base type $\tau(A)$ is completely determined by $A \setminus A_n$, the n smallest elements of A.

The bookkeeping in different versions of this definition in different attempts at a tangled web version of the proof of the consistency of NF have been different (an obvious point about the version given here is that the top ordinal α could be omitted). Another remark is that it is clear that asserting the existence of a tangled web is stronger than simple TTT: it requires $\lambda > \omega$, and the λ-completeness of course is a strong assumption in the background. All variants that I have used support versions of the following

Theorem: If there is a model of Mac Lane set theory in which there is a tangled web of cardinals τ, then NF is consistent.

Proof: Let Σ be a finite set of sentences of the language of TST. Let n be larger than any type mentioned in any formula in Σ. Partition $[\alpha]^n$ into compartments in such a way that the compartment that a set A is put into depends on the truth values of the sentences in Σ in natural models of TST_n with base type of size $\tau(B)$ where $B \setminus B_n = A$. This partition of $[\alpha]^n$ into no more than $2^{|\Sigma|}$ compartments has a homogeneous set H of size $n+1$. The natural models of TST_n with base types of size $\tau(H)$ and base types of size $\tau(H_1)$ have the same truth values for sentences in Σ, so the model of TST with base type $\tau(H)$ satisfies the restriction of the Ambiguity Scheme to Σ, so the full Ambiguity Scheme is consistent by compactness, so TST + Ambiguity is consistent so NF is consistent.

Our initial approach to proving our theorem was to attempt a Frankel-Mostowski construction of a model of Mac Lane set theory with a tangled web of cardinals. We do know how to do this, but we believe from recent experience that constructing a model of tangled type theory directly is easier, though tangled type theory is a nastier theory to describe.

We think there is merit in giving a brief description of a situation in a more familiar set theory equivalent to (a strengthening of) the very strange situation in a model of tangled type theory. This section is also useful here because it supports the discussion in the conclusion of one of the unsolved problems which is settled by this paper.

3 The model description

In this section we give a complete description of what we claim is a model of tangled type theory. Our metatheory is some fragment of ZFC.

Abstract considerations about types; the system of supertypes defined: Types are indexed by a well-ordering $\leq_\tau$ (from which we define a strict well-ordering $<_\tau$ in the obvious way). We refer to elements of the domain of $\leq_\tau$ as "type indices".

We first define a system of "supertypes" (using the same type labels).

For each element t of $\mathtt{dom}(\leq_\tau)$ we will define a set τ_t^*, called supertype t.

If m is the minimal element of the domain of $\leq_\tau$, we choose a set τ_m^* as supertype m.

$\leq_\tau$ and τ_m^* are the only parameters of the system of supertypes (which is not a model of TTT, but a sort of maximal structure for the language of TTT).

We describe the construction of τ_t^*, assuming that $t \in \mathtt{dom}(\leq_\tau)$ and $t \neq m$ and for all $u <_\tau t$, we have defined τ_u^*.

We define τ_t^* as

$$\{X \cup \{\{\tau_u^* : u <_\tau t\}\} : X \subseteq \bigcup_{u <_\tau t} \tau_u^*\}.$$

An element of supertype $t >_\tau m$ is a subset of the union of all lower types, with $t^+ = \{\tau_u^* : u <_\tau t\}$ added as an element.

Foundation in the metatheory ensures a clean construction here. An element x of supertype $t >_\tau m$ is always nonempty with t^+ as an element. The set t^+ has supertype u as an element for each $u <_\tau t$, so t^+ and so x cannot belong to any supertype u with $u <_\tau t$, by foundation: each other element of x belongs to such a supertype. We have shown that all the types are disjoint. The labelling element t^+ cannot belong to supertype t by foundation, because an element of supertype t must be nonempty and have t^+ as an element. Further, t^+ cannot belong to any supertype v with $t <_\tau v$, because any element of v contains v^+ as an element which contains supertype t as an element and any element of supertype t contains t^+ as an element, so $t^+ \in v$ would violate foundation in the metatheory.

The membership relations of this structure are transparent: $x \in_{t,u} y$ ($t <_\tau u$) is defined as $x \in \tau_t^* \wedge y \in \tau_u^* \wedge x \in y$. Considerations above show that there are no unintended memberships caused by the labelling elements t^+, because the labelling elements cannot themselves belong to any supertype. Note the presence of $\emptyset_t = \{t^+\}$ in supertype t, which has no elements of any type $u <_\tau t$ (and is distinct from $\emptyset_v$ for $v \neq t$).

The system of supertypes is certainly not a model of TTT, because it does not satisfy extensionality. It is easy to construct many sets in a higher type with the same extension over a given lower type, by modifying the other extensions of the object of higher type.

The system of supertypes does satisfy the comprehension scheme of TTT. One can use Jensen's method to construct a model of stratified comprehension with no extensionality axiom from the system of supertypes, and stratified comprehension with no extensionality axiom interprets NFU in a manner described by Marcel Crabbé in [1].

the generality of the system of supertypes: We show that any model of TTT (assuming there are any) will be in effect isomorphic to a substructure of a system of supertypes.

Let M be a model of TTT (more generally, any structure for the language of TTT in which each object is determined given all of its extensions). Let $\leq_M$ be the well-ordering on the types of M and let m be the minimal type of M. We will assume as above that $\leq_\tau$ is a well-ordering of type labels t with corresponding actual types τ_t of M: of course, we could use the actual types of M as type indices, but we preserve generality this way. We also assume that the sets implementing the types of M are disjoint (it is straightforward to transform a model in which the sets implementing the types are not disjoint to one in which they are, without disturbing its theory, by replacing each $x \in \tau_t$ with (x, t)).

We consider the supertype structure generated by $\leq_\tau := \leq_M$ and $\tau^*_m := \tau_m$. We indicate how to define an embedding from M into this supertype structure.

Define $I(x) = x$ for $x \in \tau_m = \tau^*_m$.

If we have defined I on each type $u <_\tau t$, we define $I(x)$, for $x \in \tau_t$, as $\bigcup_{u <_\tau t} \{I(y) : y \in^M_{u,t} x\} \cup \{\{\tau^*_u : u <_\tau t\}\}$.

It should be clear that as long as M satisfies the condition that an element of any type other than the base type is uniquely determined given all of its extensions in lower types, I is an isomorphism from M to a substructure of the stated system of supertypes. A model of TTT, in which any one extension of an element of any higher type in a lower type exactly determines the object of higher type, certainly satisfies this condition. So we can suppose without loss of generality that any model of TTT is a substructure of a supertype system.

Some advantages of this framework are that the membership relations in TTT are interpreted as subrelations of the membership relation of the metatheory, while the types are sensibly disjoint.

preliminaries of our construction; cardinal parameters and type -1: Now we introduce the notions of our particular construction in this framework.

Let λ be a limit ordinal. Let $\leq_\tau$ be the order on $\lambda \cup \{-1\}$ which has -1 as minimal and agrees otherwise with the usual order on λ.

Let $\kappa > \lambda$ be a regular uncountable ordinal. Sets of cardinality $< \kappa$ we call "small". Sets which are not small we may call "large".

Let μ be a strong limit cardinal greater than κ with cofinality at least κ.

Let $\tau^*_{-1} = \tau_{-1}$ be

$$\{(\nu, \beta, \gamma, \alpha) : \nu < \mu \wedge \beta \in \lambda \cup \{-1\} \wedge \gamma \in \lambda \setminus \{\beta\} \wedge \alpha < \kappa\}.$$

Note that this completes the definition of the supertype structure we are working in: we now have a definite reference for τ^*_α for $\alpha \in \lambda$.

type shorthand: Notice that if α, β are types, $\alpha \in \beta$ is a convenient short way to say $-1 <_\tau \alpha <_\tau \beta$. We will usually write $<$ instead of $<_\tau$. Notice that if $\alpha <_\tau \beta$ are types, and $x \in \tau_\alpha$, $x \cap \tau_\beta$ is the extension of x over supertype β: we presume here that $\tau_\gamma \subseteq \tau^*_\gamma$ for each type γ.

A nonempty finite subset of $\lambda \cup \{-1\}$ may be termed an *extended type index.* If A is an extended type index with at least two elements, A_1 is defined as $A \setminus \{\mathtt{min}(A)\}$.

atoms, litters and near-litters: We may refer to elements of τ_{-1} (or closely related objects) as "atoms" from time to time, though they are certainly not atomic in terms of the metatheory.

A *litter* is a subset of τ_{-1} of the form $L_{\nu,\beta,\gamma} = \{(\nu, \beta, \gamma, \alpha) : \alpha < \kappa\}$. The litters make up a partition of type -1 (which is of size μ) into size κ sets.

On each litter $L = L_{\nu,\beta,\gamma}$ define a well-ordering $\leq_L$: $(\nu, \beta, \gamma, \alpha) \leq_L (\nu, \beta, \gamma, \alpha')$ iff $\alpha \leq \alpha'$. The strict well-ordering $<_L$ is defined in the obvious way.

A *near-litter* is a subset of τ_{-1} with small symmetric difference from a litter. We define $M \sim N$ as $|M \Delta N| < \kappa$, for M, N near-litters: in English, we say "M is near N" iff $M \sim N$. Note that nearness is an equivalence relation on near-litters. Note that there are μ near-litters, because the cofinality of μ is at least κ.

We define N°, for N a near-litter, as the litter L such that $L \sim N$.

Define $\mathbf{X}_{\beta,\gamma}$ as $\{L_{\nu,\beta,\gamma} : \nu < \mu\}$. This gives us a set of litters of size μ for each pair of types $\beta \in \lambda \cup \{-1\}$ and $\gamma \in \lambda \setminus \{\beta\}$: the collection of sets $\mathbf{X}_{\beta,\gamma}$ is pairwise disjoint.

enforcing extensionality in the type system: We describe the mechanism which enforces extensionality in the substructure of this supertype structure that we will build.

The levels of the structure we will define are denoted by τ_α for $\alpha \in \lambda \cup \{-1\}$. As we have already noted, $\tau_{-1} = \tau^*_{-1}$ as defined above.

In defining $\tau_\alpha \subseteq \tau^*_\alpha$ for each α, we assume that we have already defined τ_β for each $\beta < \alpha$, and that the system of types $\{\tau_\beta : \beta <_\tau \alpha\}$ already defined satisfies various hypotheses which we will discuss as we go. Elements x

of τ_α^* which we consider for membership in τ_α will have $x \cap \tau_\beta^* \subseteq \tau_\beta$ for $\beta < \alpha$. We assume that for $\gamma < \beta < \alpha$, if $x \in \tau_\beta$, $x \cap \tau_\gamma^* \subseteq \tau_\gamma$.

We suppose that each τ_β already constructed is of cardinality μ. Note that we already know that τ_{-1} is of cardinality μ.

We postulate a well-ordering $\leq_\beta$ of τ_β, of order type μ, with associated strict well-ordering $<_\beta$, for each $\beta < \alpha$. Note that these well-orderings do not depend on α: once constructed, they remain the same at later stages. Some conditions on these well-orderings will be stated later. We define $\iota_*(x)$, where $x \in \tau_\gamma$, $[\gamma < \alpha]$ as the order type of the restriction of $\leq_\gamma$ to $\{y : y <_\gamma x\}$. Note that this gives us the ability to compare the ordinal position of elements of different types.

We further intimate that for each $x \in \tau_\gamma$, $-1 < \gamma < \alpha$, we have defined objects S for which we say that S is a support of x. The definition of supports will be given later. For the moment, we define τ_γ^+ as the set of all (x, S) for which $x \in \tau_\gamma$ and S is a support of x. It is a hypothesis of the recursion that τ_γ^+ is of cardinality μ, and we also provide a well-ordering $\leq_\gamma^+$ of order type μ of τ_γ^+ . We provide that τ_{-1}^+ is the set of all $(x, \emptyset)$ for $x \in \tau_{-1}$.

We provide that for every near litter N and every $\beta < \alpha$, there is a unique element N_β of τ_β such that $N_\beta \cap \tau_{-1} = N$ (we will quite shortly give a precise description of all extensions of this object).

Definition: If X is a subset of type γ and $\gamma < \beta$, we define X_β as the unique element Y of τ_β such that $Y \cap \tau_\gamma = X$ (if this exists). Of course, this notation is only usable to the extent that we suppose that extensionality holds. Notice that the notation N_β is a case of this.

If $x \in \tau_{-1}$, we refer to any $\{x\}_\beta$ as a *typed atom* (of type β) and if N is a near-litter we refer to N_β as a *typed near-litter* (of type β).

We further stipulate that extensionality holds for each $\beta \in \alpha$ (for each $\beta \in \gamma$ and $\gamma \in \beta$, any $x \in \tau_\beta$ is uniquely determined by $x \cap \tau_\gamma$; x is uniquely determined by $x \cap \tau_{-1}$ only on the additional assumption that $x \cap \tau_{-1}$ is nonempty)[NOTE: I do not know whether this condition is still in use: , and that for any $\beta \in \alpha$ and $x \in \tau_{-1}$, $\iota_*(\{x\}_\beta) = \iota_*(\{x\}_0)$].

We define for each $\beta, \gamma < \alpha$ a function $f_{\beta,\gamma}$ (whose definition does not actually depend on α: it will be the same at every stage). $f_{\beta,\gamma}$ is an injection from τ_β^+ into $\mathbf{X}_{\beta,\gamma}$. When we define $f_{\beta,\gamma}(x)$, we presume that we have already defined it for $y <_\beta^+ x$. We define $f_{\beta,\gamma}(x)$ as $L \cap \tau_{-1}$, where L is $<_\gamma$-first such that $L \cap \tau_{-1} \in \mathbf{X}_{\beta,\gamma}$, and for every $N \sim L \cap \tau_{-1}$, $\iota_*(N_\gamma) > \iota_*(\pi_1(x))$, and for any $y <_\beta^+ x$, $f_{\beta,\gamma}(y) \neq L \cap \tau_{-1}$.

We define the notion of *pre-extensional* element of τ_β^* ($\beta \leq \alpha$). An element x of τ_β^* is pre-extensional iff there is a $\gamma < \beta$ such that (1) $x \cap \tau_\gamma^* \subseteq \tau_\gamma$, and (2) $\gamma = -1$ if $x \cap \tau_{-1}$ is nonempty or if any $x \cap \tau_\delta$ ($\delta <_\tau \beta$) is empty, and (3) for each $\delta \in \beta \setminus \{\gamma\}$,

$$x \cap \tau_\delta = \{N_\delta : (\exists a \in x \cap \tau_\gamma : \exists S : N \sim f_{\gamma,\delta}(a, S))\}.$$

We say for any $x \in \tau^*_\beta$ and γ with this property that $x \cap \tau_\gamma$ is a distinguished extension of x.

We presume that all elements of τ_β, $\beta < \alpha$, are pre-extensional.

Note that we now know how to compute all other extensions of typed near-litters N_β, because the -1-extension of N_β is distinguished and this indicates how to compute all the other extensions.

We note that the order conditions in the definition of $f_{\beta,\gamma}$ ensure that for any $x \in \tau_\beta$ there is only one $\gamma <_\tau \beta$ such that $x \cap \tau_\gamma$ is a distinguished extension of x. If any $x \cap \tau_\gamma$ is empty or if $x \cap \tau_{-1}$ is not empty, $\gamma = -1$ is the unique possibility. So, what remains is the case of x with $x \cap \tau_{-1}$ empty and each $x \cap \tau_\gamma$ nonempty for $\gamma < \beta$. Now the order conditions on $f_{\beta,\gamma}$ ensure that the following procedure works: if $x \cap \tau_\gamma$ is a distinguished extension, we can choose $a \in x \cap \tau_\gamma$ with $\iota_*(a)$ minimal. Any $c \in x \cap \tau_\delta$ ($\delta \neq \gamma$) then is an N_δ and has $N \sim f_{\gamma,\delta}(b, S)$ for some $b \in \tau_\gamma$ and support S of b, and we have $\iota_*(a) \leq \iota_*(b) < \iota_*(N_\delta) = \iota_*(c)$, so the distinguished extension can be identified by finding the smallest value of ι_* on an extension.

For any $\delta \in \alpha$ and nonempty subset a of type $\gamma \neq \delta$, we define $A_\delta(a)$ as

$$\{N_\delta : (\exists x \in a : \exists S : N \sim f_{\gamma,\delta}(x, S))\}.$$

For any nonempty subset of type δ there is at most one subset y of any type such that $A_\delta(y) = x$. There cannot be more than one such y in any given type because the f maps are injective. There cannot be more than one such y in different types because the ranges of f maps with distinct index pairs are disjoint. We use the notation $A^{-1}(y)$ for this set if it exists, defining a very partial function A^{-1} on subsets of types.

Note that the distinguished extension of any type element x is the image under A^{-1} of the other extensions.

Let a be a subset of type δ for which $A^{-1}(a)$ exists and is a subset of type γ. We argue that the minimum of ι_* on $A^{-1}(a)$ is less than the minimum of ι_* on a (this is basically the same as an argument given above but we give it for completeness). Let $b \in a$ have $\iota_*(b)$ minimal. Let $c \in A^{-1}(a)$ have $\iota_*(c)$ minimal. We will have $b = N_\delta$ for $N \sim f_{\gamma,\delta}(d, S)$ for some $d \in A^{-1}(a)$ and support S of d. We then have $\iota_*(c) \leq \iota_*(d) < \iota_*(b)$ by the order requirements in the definition of the f maps.

It follows that no subset of a type has infinitely many iterated images under A^{-1}.

We say that an element of a type is *extensional* iff it is pre-extensional and its distinguished extension has an even number of iterated images under

A^{-1}. This implies that each of its other extensions has an odd number of iterated images under A^{-1}. This is enough to ensure that two extensional model elements with any common extension will be equal: if two extensional model elements have an empty extension in common, they both have all extensions empty and are equal. If two extensional model elements have a nonempty extension in common, it will be the distinguished extension of both, or a non-distinguished extension of both, since distinguished and non-distinguished extensions are taken from disjoint classes of subsets of types (when nonempty). In either case we deduce that both have the same distinguished extension and thus have all extensions the same and are equal. Note that this gives weak extensionality over τ_{-1} (many objects have empty extension over type -1) but it gives full extensionality over any other type.

We introduce the notation (β, δ, D) where $\delta < \beta$ and $D \subseteq \tau_\delta$. This stands for the unique extensional element x of τ^*_β such that $x \cap \tau_\delta = D$. It should be clear that there is only one such object. If D is empty, it is the unique element of τ^*_β with empty intersection with each τ^*_γ for $\gamma < \beta$. If $\delta = -1$ and D is nonempty, or if $\delta > -1$ and D has an even number of iterated images under A^{-1}, then it is the unique element x of of τ^*_β which is extensional and has distinguished extension $x \cap \tau_\delta$. If D is nonempty and has an odd number of iterated images under A^{-1}, let $A^{-1}(D) \subseteq \tau_\gamma$, and it is the same as $(\beta, \gamma, A^{-1}(D))$. This notation is mainly for compatibility with previous versions of the paper, but may have its uses.

We assume that all elements of τ_β's already constructed are extensional.

brief note on our further needs: A crucial aspect of this is that we will need to define τ^+_α so that it has cardinality μ for the process to continue. It is certainly not a sufficient restriction to require elements of τ_α to be extensional: we will require a further symmetry condition.

structural permutations defined: We define classes of permutations of our structures.

A -1-structural permutation is a permutation of τ_{-1}.

An α-structural permutation is a permutation π of τ_α such that for each type $\beta < \alpha$ there is a β-structural permutation π_β such that $\pi(x) \cap \tau_\beta = \pi_\beta$“$(x \cap \tau_\beta)$ for any $x \in \tau_\beta$.

derivatives of structural permutations: The maps π_β are referred to as derivatives of π. More generally, for any finite subset A of $\lambda \cup \{-1\}$ with maximum α, define π_A as $(\pi_{A \setminus \{\mathtt{min}(A)\}})_{\mathtt{min}(A)}$. The maps π_A may be referred to as iterated derivatives of π. It should be clear that a structural permutation is exactly determined by its iterated derivatives which are -1-structural.

allowable permutations defined: Structural permutations are defined on the supertype structure generally. We need a subclass of structural permutations which respects our extensionality requirements.

A -1-allowable permutation is a permutation π of τ_{-1} such that for any near-litter N, π"N is a near-litter.

An α-allowable permutation is an α-structural permutation, each of whose derivatives π_β is a β-allowable permutation (and satisfies the condition that π_β"$\tau_\beta = \tau_\beta$) and which satisfies a coherence condition relating the f maps and derivatives of the permutation: for suitable $\beta, \gamma < \alpha$,

$$f_{\beta,\gamma}(\pi_\beta(x), \pi_\beta[S]) \sim (\pi_\gamma)_{-1}\text{“}f_{\beta,\gamma}(x, S).$$

(where the action of structural permutations on supports will be defined shortly).

Note that an α-allowable permutation is actually defined on the entire supertype structure, though what interests us about it is its actions on objects in our purported TTT model.

supports defined: Where $0 \leq \beta \leq \alpha$, a β*-support* is defined as a well-ordering $\leq_S$ of pairs (x, A), where A is a finite subset of λ with minimum $\gamma \geq 0$ and maximum β and $x \in \tau_\gamma$ has $x \cap \tau_{-1}$ either a singleton or a near-litter. By fiat, we define $\emptyset$ as the only -1-support.

We define the action of a β-allowable permutation π on a β-support $\leq$: $\pi[\leq_S] = \{((\pi_A(x), A), (\pi_B(y), B))) : (x, A) <_S (y, B)\}$. In the case of -1-supports, $\pi[\emptyset] = \emptyset$.

An element x of τ_β^* ($\beta \geq 0$) has β-support $\leq_S$ iff for every β-allowable permutation π, if $\pi[\leq_S] = \leq_S$ then $\pi(x) = x$. We say that an element x of τ_β^* which has a β-support is β-symmetric. Every element of τ_{-1} has $\emptyset$ as its only support [by technical fiat].

It is straightforward to observe that there are μ supports, since there are μ atoms, μ near-litters (this depends on the cofinality of μ being at least κ) and $< \kappa$ finite sets of elements of λ. Thus τ_β^+ is already known to be of size μ for $\beta < \alpha$.

It is important to note that if $\leq_S$ is a support of $x \in \tau_\beta$, $\pi[\leq_S]$ is a support of $\pi(x)$ for any β-allowable permutation π.

We may sometimes refer to a support $\leq_S$ as simply S (when we do not explicitly use it as an order). We will use the notation S_γ for the element (x, A) of $\mathtt{dom}(S)$ such that the order of the restriction of $<_S$ to $\{(y, B) \in \mathtt{dom}(S) : (y, B) \leq_S (x, A)\}$ is γ.

motivation of the coherence condition: The motivation for this is that we need α-allowable permutations to send extensional elements of supertypes to extensional elements. Suppose $x \cap \tau_\beta = \{b\}$. If x is extensional, this has to be the distinguished extension of x. For any $\gamma \in \alpha \setminus \{\beta\}$, it follows that $x \cap \tau_\gamma$ is the set of all N_γ such that $N \sim f_{\beta,\gamma}(b, S)$ for some support S of b. This tells us that an α-allowable permutation π, for which we must have that $\pi(x)$ has β-extension $\{\pi_\beta(b)\}$, must have the γ-extension of $\pi(x)$ equal to π_γ"$\{N_\delta : \exists S : N \sim f_{\beta,\gamma}(b, S)\}$ but must also have its

γ-extension equal to $\{N_\delta : \exists S : N \sim f_{\beta,\gamma}(\pi_\beta(b), S)\}$. This tells us that $\pi_\gamma(f_{\beta,\gamma}(b,S)_\delta) \in \{N_\delta : (\exists T : N \sim f_{\beta,\gamma}(\pi_\beta(b), T)\}$ for each support S of b. The coherence condition enforces this neatly, showing that it is motivated by considerations required to get extensionality to work: the action of π_β conveniently correlates supports of b with supports of $\pi_\beta(b)$.

We defined $A_\delta(a)$ as

$$\{N_\delta : (\exists x \in a : (\exists S : N \sim f_{\gamma,\delta}(a)))\}.$$

If π is allowable of suitable index, π_δ"$A_\delta(a) = A_\delta(\pi_\gamma$"$a)$ follows from the coherence condition. Verify this:

Suppose we have N_δ with $x \in a$ such that $N \sim f_{\gamma,\delta}(x, S)$. Then

$$\pi_\delta(N_\delta) \cap \tau_{-1} = (\pi_\delta)_{-1}``N \sim (\pi_\delta)_{-1}``f_{\gamma,\delta}(x,S) \sim f_{\gamma,\delta}(\pi_\gamma(x), \pi_\gamma[S]).$$

So any element of π_δ"$A_\delta(a)$ is in $A_\delta(\pi_\gamma$"$a)$.

Suppose we have N_δ with $x \in a$ such that $N \sim f_{\gamma,\delta}(\pi_\gamma(x), S)$. We then have $N \sim (\pi_\delta)_{-1}``f_{\gamma,\delta}(x, \pi_\gamma^{-1}[S])$. We want to show that $\pi_\delta^{-1}(N_\delta) \in A_\delta(a)$. $\pi_\delta^{-1}(N_\delta) \cap \tau_{-1} = (\pi_\delta)_{-1}^{-1}``N \sim (\pi_\delta)_{-1}^{-1}``((\pi_\delta)_{-1}``f_{\gamma,\delta}(x, \pi_\gamma^{-1}[S])) = f_{\gamma,\delta}(x, \pi_\gamma^{-1}[S])$, establishing what we need.

Notice that this shows that the coherence condition implies that the image under an allowable permutation of a pre-extensional element of our structure is pre-extensional.

Now this implies that if $a \subseteq \tau_\gamma$, then $A^{-1}(a)$ exists and is in τ_δ exactly if $A^{-1}(\pi_\gamma$"$a)$ exists and is in τ_δ, and moreover $A^{-1}(\pi_\gamma$"$a)$ is equal to π_δ"$A^{-1}(a)$ if it exists under these conditions. This verifies that the coherence condition implies that allowable permutations preserve full extensionality, not just pre-extensionality: the number of iterated images under A^{-1} of an extension that exist is not affected by application of an allowable permutation in a suitable sense.

the definition of τ_α: We stipulate that all elements of τ_β have β-supports, and define τ_α as the set of elements x of τ_α^* such that $x \cap \tau_\beta^* \subseteq \tau_\beta$ for each $\beta < \alpha$, x is extensional, and x has an α-support.

We still have to prove that the cardinality of τ_α, and so of τ_α^+, is μ, to show that the construction works.

Observation (κ-completeness of the structure): For any subset X with cardinality $< \kappa$ of a type γ and $\beta > \gamma$, it should be clear that X_β has a support, computable from the union of the supports of the elements of X by replacing each element (u, B) of the union with $(u, B \cup \{\beta\})$, and therefore belongs to the model. X_β is obviously extensional (the extension X is clearly the distinguished extension and has no image under A^{-1}).

conditions on choice of the distinguished well-orderings of types: The well-ordering $<_\alpha$ of τ_α can be chosen freely.

The well-ordering $<^+_\alpha$ of τ^+_α must satisfy the condition that for each $(x, S) \in \tau^+_\alpha$, for each $(z, A) \in S$ such that $f_{\beta,\gamma}(y, T) = L$ for some $-1 < \beta < \mathtt{min}(A_1), \gamma = \mathtt{min}(A) < \alpha, y, T$, for L the litter either including the singleton $z \cap \tau_{-1}$ or near the near-litter $z \cap \tau_{-1}$, $\iota^+_*(y, T) <^+_\alpha \iota^+_*(x, S)$ [where ι^+_* computes ordinal position in any τ^+_β.] If $f_{-1,\gamma}(y, T) = L$, we instead require that $\iota^+_*(M_{\mathtt{min}(A)}, \{(M_{\mathtt{min}(A)}, A))\}) <^+_\alpha (x, S)$, where M is the litter containing y. There is no obstruction to enforcing this condition, because there are many elements of τ^+_α for which there is no such obligation, which can be used to fill gaps, as it were.

It should be noted that type 0 has a very simple description: the -1-extensions of type 0 objects are exactly the sets with small symmetric difference from small or co-small unions of litters, and that these are the same extensions over type -1 which appear in any positive type.

At this point we have a complete description of the structure which we claim is a model of TTT.

4 Verification that the structure defined is a model

4.1 The freedom of action theorem

Definition (approximation): A β-approximation is a map π^0 from finite subsets of λ with maximum element β such that each $\pi^0(A)$ (which we write π^0_A) is a function with the following properties:

1. The domain and image of π^0_A are the same and π^0_A is injective.
2. Each domain element x of π^0_A is such that (x, A) is a condition and $x \cap \tau_{-1}$ is included in a typed litter (this excludes some typed near-litters as values of x).
3. $x \cap \tau_{-1}$ and $\pi^0_A(x) \cap \tau_{-1}$ have the same cardinality, which is either 1 or κ, since the previous condition tells us that x is a typed atom or near-litter.
4. For each A, the collection $\{x \cap \tau_{-1} : x \in \mathtt{dom}(\pi^0_A)\}$ is pairwise disjoint and a small subcollection of this set covers any litter with which its union has large (i.e, cardinality κ) intersection.

We say that π^0 approximates a β-allowable permutation π just in case $\pi_A(x) = \pi^0_A(x)$ whenever the latter is defined.

litter near a near-litter: For any near-litter N, define N° as the unique litter L such that $L \sim N$.

Definition (flexibility): A typed near-litter x is A-flexible if it is of type $\mathtt{min}(A)$ and $(x \cap \tau_{-1})^\circ$ is not in the range of any $f_{\gamma,\mathtt{min}(A)}$ for $\gamma < \mathtt{min}(A_1)$.

Definition (exception, exact approximation): A -1-allowable permutation π has *exception* x if, L being the litter containing x, we have either $\pi(x) \notin (\pi``L)^\circ$ or $\pi^{-1}(x) \notin (\pi^{-1}``L)^\circ$.

A β-approximation π^0 *exactly approximates* a β-allowable permutation π iff π^0 approximates π and for every exception x of a $\pi_{A\cup\{-1\}}$ (A not containing -1) we have $\{x\}_{\mathtt{min}(A)}$ in the domain of π^0_A.

Theorem (freedom of action): A β-approximation π^0 will exactly approximate some β-allowable permutation π if it satisfies the additional condition that any domain element x of π^0_A which is a typed near-litter is A-flexible.

Proof: For each pair of sets L, M which are co-small subsets of litters, we define $\pi_{L,M}$ as the unique map ρ from L onto M such that for any $x, y \in L$,

$$x <_{L^\circ} y \leftrightarrow \rho(x) <_{M^\circ} \rho(y):$$

$\pi_{L,M}$ is the unique map from L onto M which is strictly increasing in the order determined by fourth projections of atoms. Notice that if $L' \subseteq M$ and $M' = \pi_{L,M}``L'$, then $\pi_{L',M'} \subseteq \pi_{L,M}$. [NOTE: the choice of these maps does not need to be so concrete, but the fact that it can be indicates for example that there is no use of choice here.]

We also choose an extension of each π^0_A to suitable near-litter subsets of all A-flexible typed litters (the conditions dictate what near-litters should be added to the domain, uniquely, since the litters need to be exactly covered by domain elements in a suitable sense); we do this without notational comment, simply assuming that π^0_A is defined for each A and A-flexible litter M at some near-litter subset of M, which can be arranged harmlessly (for example, one could have π^0_A act as the identity on the new A-flexible typed near-litters, but we do not require this).

We choose an approximation π^0 satisfying the conditions of the theorem and extend it as indicated in the previous paragraph. We compute the allowable permutation π on all support conditions (and therefore compute all its derivatives π_A at all atoms ($-1 \in A$), so completely defining it, using the assumption that we already know how to carry out this construction to define γ-allowable permutations exactly approximated by any given γ-approximation for $\gamma < \beta$.

We use the notation π^*_A for the partially computed π_A at any point in the calculation before we are done.

We first indicate how to compute $\pi^*_A(L_{\mathtt{min}(A)})$, where L is a litter.

If L is A-flexible, we can compute $\pi^*_A(L_{\mathtt{min}(A)})$ as the union of all $\pi^0_A(M)$ for $M \subseteq L$.

We further extend this to describe the action of $\pi_{A\cup\{-1\}}$ on elements of L: for each $x \in L$, if $\{x\}_{\mathtt{min}(A)}$ is in the domain of π^0_A, which maps it to

$\{y\}_{\mathtt{min}(A)}$, $\pi_{A\cup\{-1\}}$ maps x to y. Define L^- as the set of all $x \in L$ such that $\{x\}_{\mathtt{min}(A)}$ is not in the domain of π^0_A and define M^- as the set of all $x \in M$ such that $\{x\}_{\mathtt{min}(A)}$ is not in the domain of π^0_A (notice that $L^-_{\mathtt{min}(A)}$ and $M^-_{\mathtt{min}(A)}$ are in the domain of π^0_A.). For $x \in L^-$ we define $\pi^*_{A\cup\{-1\}(x)}$ as $\pi_{L^-,M^-}(x)$.

If L is A-inflexible, we have $f_{\gamma,\mathtt{min}(A)}(x,S) = L$ for some $\gamma < \mathtt{min}(A_1)$, $x \in \tau_\gamma$, and S a support of x.

We expect $\pi^*_A(f_{\gamma,\mathtt{min}(A)}(x,S)_{\mathtt{min}(A)})$ to be near

$$f_{\gamma,\mathtt{min}(A)}(\pi^*_{A_1\cup\{\gamma\}}(x), \pi^*_{A_1\cup\{\gamma\}}[S]))_{\mathtt{min}(A)}.$$

The hypothesis of the recursion allows us to compute values of $\pi^*_{A_1\cup\{\gamma\}}$ and images under this permutation of γ-supports, because we can consider a γ-approximation ρ^0 such that $\rho^0_B(y) = \pi^*_{A_1\cup B}(y)$ when y is a suitable typed near-litter and the latter is defined, using the additional hypothesis that we have computed $\pi^*_{A_1\cup C}(z)$ where $\mathtt{max}(C) = \gamma$ and z is any item of appropriate type which a subset of the image under an f map of something with smaller image under ι^+_* than (x,S), which will cover all support conditions in S, otherwise making arbitrary assignments of values to B-flexible items for which we have not computed a value. The approximation ρ_0 has as singleton elements in its domain exactly those justified by reference to π_0: $\rho^0_B(w) = \pi^0_{A_1\cup B}(w)$ for appropriate typed atoms w [this allows us to determine exactly what typed near-litters are in the domain of ρ^0]. By inductive hypothesis, we can determine the allowable permutation ρ exactly approximating ρ_0 and use ρ_C to compute $\pi^*_{A_1\cup C}$ for C with maximum γ at x as required for x and for items in S, supporting the computation of the expression given. ρ_C will not necessarily agree with the eventually constructed $\pi_{A_1\cup C}$ everywhere, but appropriate maps will agree at x because the actions on the support S will be correct.

In this way we have computed $\pi^*_A(L)^\circ$, which we call M. Let L^- be the largest subset of L which does not include any x such that $\{x\}_\alpha$ is in the domain of π^0_A. Let M^- be the largest subset of M which does not include any x such that $\{x\}_\alpha$ is in the domain of π^0_A. We can define $\pi^*_A(L_{\mathtt{min}(A)})$ as the union of $M^-_{\mathtt{min}(A)}$ and all $\pi^*_A(\{x\}_{\mathtt{min}(A)})$ for $x \in L \setminus L^-$.

We further extend this to describe the action of $\pi_{A\cup\{-1\}}$ on elements of L: for each $x \in L$, if $\{x\}_{\mathtt{min}(A)}$ is in the domain of π^0_A, which maps it to $\{y\}_{\mathtt{min}(A)}$, $\pi_{A\cup\{-1\}}$ maps x to y. For $x \in L^-$ we define $\pi^*_A(x)$ as $\pi_{L^-,M^-}(x)$.

The process given will compute $\pi_A(x)$ for every atom x. Since the action on every atom is fixed, π is fixed as a structural permutation.

The method by which the derivatives of π are evaluated at atoms ensures that π_A agrees with π^0_A on typed singletons. It also ensures that (if π and its derivatives defined as indicated satisfy the coherence conditions) $\pi_{A\cup\{-1\}}$ has an exception x only if $\{x\}_{\mathtt{min}(A)}$ is in the domain of π^0_A.

The method of computation verifies that the coherence conditions will hold. The method of computation also verifies that π is a permutation, as π^{-1} is computed in precisely the same way from $(\pi^0)^{-1}$.

4.2 Types are of size μ (so the construction actually succeeds)

Now we argue that (given that everything worked out correctly already at lower types) each type α is of size μ, which ensures that the construction actually succeeds at every type.

Definition (coding functions): For any support S and object x, we can define a function $\chi_{x,S}$ which sends $T = \pi[S]$ to $\pi(x)$ for every T in the orbit of S under the action of allowable permutations. We call such functions *coding functions*. Note that if $\pi[S] = \pi'[S]$ then $(\pi^{-1} \circ \pi')[S] = S$, so $(\pi^{-1} \circ \pi')(x) = x$, so $\pi(x) = \pi'(x)$, ensuring that the map $\chi_{x,S}$ for which we gave an implicit definition is well defined.

Definition (the specification of a support): For each support S we define a combinatorial object S^* which we call its *specification*. We will show below that what it specifies is the orbit in the action of allowable permutations on supports to which it belongs.

For S a support, we define S_ϵ as the element x of its domain such that the restriction of S to $\{y : y <_S x\}$ is of order type ϵ.

The specification S^* is a well-ordering of the same length as S. We describe the elements of its domain.

1. If S_ϵ is $(\{x\}_\beta, A)$, then S^*_ϵ is $(0, \beta, \delta, \epsilon, A)$ such that S_δ is (N_β, A), for N a near-litter with x belonging to N, or $(0, \beta, \kappa, \epsilon, A)$ if there is no such (N_β, A). There is at most one such δ since typed near-litters in supports are disjoint.

   One can state an internal condition on specifications that if S^*_ϵ is $(0, \beta, \delta, \epsilon, A)$, then S^*_δ must have first component 2 or 3.

2. If S_ϵ is (N_β, A) and N is a near-litter, and either $|A| = 1$ or N° is not in the range of any $f_{\gamma,\beta}$ for $\gamma < \mathtt{min}(A_1)$, then S^*_ϵ is $(1, \beta, \epsilon, A)$.

3. If S_ϵ is (N_β, A) and N is a near-litter, and $N^\circ = f_{\gamma,\beta}(x, T)$ for $-1 < \gamma < \mathtt{min}(A_1)$, for $x \in \tau_\gamma$ then S^*_ϵ is $(2, \beta, \chi_{x,T}, \epsilon, A)$: the third component is the coding function with domain T which yields x.

4. If S_ϵ is (N_β, A) and N is a near-litter, and $N^\circ = f_{-1,\beta}(x, \emptyset)$ then S^*_ϵ is $(3, \beta, \delta, \epsilon, A)$, where S_δ is $(\{x\}_{\mathtt{min}(A_1)}, \mathtt{min}(A_1))$, or $(3, \beta, \kappa, \epsilon, A)$, if there is no such δ.

Observation: On the inductive hypothesis that there $< \mu$ γ-coding functions for each $\gamma < \alpha$, we observe that there are $< \mu$ specifications of β-supports for $\beta \leq \alpha$.

Lemma: The specification of a β-support exactly determines the orbit in the action of β-allowable permutations on supports to which it belongs.

Proof of Lemma: It is straightforward to see that if S is a β-support and if π is a β-allowable permutation, that $(\pi[S])^* = S^*$. The relationships between items in the support recorded in the specification are invariant under application of allowable permutations.

It remains to show that if S and T are supports, and $S^* = T^*$, there is an allowable permutation ϕ such that $\pi[S] = T$.

We construct π using the Freedom of Action Theorem.

If we have $S_\epsilon = (\{x\}_\beta, A)$, we will have $T_\epsilon = (\{y\}_\beta, A)$ for some y, and we will set $\pi^0_A(\{x\}_\beta) = \{y\}_\beta$ as part of the construction of the local bijection to be used.

If we have $S_\epsilon = ((M_\beta, A)$ for M a near litter and either $|A| = 1$ or M° is not in the range of any $f_{\gamma,\beta}$ for $\gamma < \texttt{min}(A_1)$, then $T_\epsilon = (N_\beta, A)$ for N a near litter, with analogous properties, and we set $\pi^0_A(M^\circ_\beta) = N^\circ_\beta$ as part of the data for application of the Freedom of Action Theorem [actually, we set $\pi^0_A((M^-_\beta) = N^-_\beta$ for suitable subsets of M^- and N^-, excluding any atoms in M or N that are assigned values; this adjustment can be made at the end of the process].

If we have $S_\epsilon = (M_\beta, A)$ for M a near litter with $M^\circ = f_{\gamma,\beta}(x, T)$, where $-1 < \gamma < \texttt{min}(A_1)$, then S^*_ϵ is $(2, \beta, \chi_{x,T}, A)$ and where $T_\epsilon = (N_\beta, A)$, T^*_ϵ is $(2, \beta, \chi_{y,T'}, A)$,, where $N^\circ = f_{\gamma,\beta}(y, T')$ and the coding functions $\chi_{x,T}$ and $\chi_{y,T'}$ are the same.

T' is $\pi_0[T]$ for some allowable permutation, and $T_\epsilon = (N_\beta, A)$ with $N \sim \pi_0(M_\beta) \cap \tau_{-1}$, by the definition of coding functions. [NOTE: add discussion of how the approximation is augmented – basically, map items in T to corresponding items in T']

and this implies that $\pi(M_\beta) \cap \tau_{-1} \Delta N_\beta \cap \tau_{-1}$ is small: we need to augment the local bijection to prevent anomalies, and there is a way to do this.

We want to ensure that elements of $M \setminus M^\circ$ and elements of $M^\circ \setminus M$ correlate with typed atoms in the domain of π^0_A and sent to elements of N_β, and similarly elements of M are chosen to have their associated typed atoms mapped by π^0_A to type correlates of elements of $N \setminus N^\circ$ and $N^\circ \setminus N$. Some additional work must be done. For each new element introduced to the domain of π^0_A, we have the obligation to fill in its complete orbit in π^0_A. The restriction we must obey as we do this is that any element of a near-litter whose typed version appears with index A in S must be mapped by $(\pi^0)_{A\cup\{-1\}}$ to an element of the corresponding near-litter in T and any element of a near-litter in T whose typed version appears with index in A in S must be mapped by $((\pi^0)^{-1}_{A\cup\{-1\}}$ to an element of the corresponding near-litter in S. Since only countably many new values are needed to fill in each orbit and κ is uncountable, there is no obstruction to doing this. Note that atoms already in the domain of π^0_A are already constrained to behave in this way. The map eventually constructed by Freedom of Action will send M_β to N_β because it maps typed atoms correlated with elements of $M\Delta M^\circ$ to and elements of $N\Delta N^\circ$ to appropriate values individually, and all other typed atoms "in" M_β must be mapped to values in N_β (and typed atom "elements" of N_β mapped from elements of M_β) because the map constructed by Freedom of Action has no exceptions not in the domain of the local bijection.

If we have $S_\epsilon = (M_\beta, A)$ for M a near litter with $M^\circ = f_{-1,\beta}(x, \emptyset)$, then we have $T_\epsilon = (N_\beta, A)$ where $N^\circ = f_{-1,\beta}(y, \emptyset)$, and we add to our approximation the information that $\{x\}_{\mathtt{min}(A_1)}$ is mapped to $\{y\}_{\mathtt{min}(A_1)}$ by $\pi^0_{A_1}$, which enforces mapping of M_β to N_β up to nearness, and the fix to get M_β to map precisely to N_β is as in the previous case.

So we have completed the description of what we need to do to construct the needed permutation.

Since the specifications precisely determine the orbits in supports under allowable permutations, and there are $< \mu$ specifications (on stated hypotheses) there are $< \mu$ such orbits.

The strategy of our argument for the size of the types is to show that that there are $< \mu$ coding functions whose domain includes a strong support for each type , which implies that there are no more than μ (and so exactly μ) elements of each type, since every element of a type is obtainable by applying a coding function (of which there are $< \mu$) to a support (of which there are μ), and every element of a type has a strong support.

Analysis of coding functions for type 0: We describe all coding functions for type 0 (without concerning ourselves about whether supports are strong). The orbit of a 0-support in the allowable permutations is determined by the positions in the support order occupied by near-litters, and for each position in the support order occupied by a singleton, the position, if any, of the near-litter in the support order which includes it. There are no more than 2^κ ways to specify an orbit. Now for each such equivalence class, there is a natural partition of type -1 into near-litters, singletons, and a large complement set. Notice that near-litters in the partition will be obtained by removing any singletons in the domain of the support which are included in them. The partition has $\nu < \kappa$ elements, and there will be $2^\nu \leq 2^\kappa$ coding functions for that orbit in the supports, determined by specifying for each compartment in the partition whether it is to be included or excluded from the set computed from a support in that orbit. So there are no more than $2^\kappa < \mu$ coding functions over type 0.

Analysis of the general case:

Our inductive hypothesis is that for each $\beta < \alpha$ we have $< \mu$ β-coding functions.

We specify an object $X \in \tau_\alpha$ and an α-support S for X, and develop a recipe for the coding function $\chi_{X,S}$ which can be used to see that there are $< \mu$ α-coding functions (assuming of course that we know that things worked out correctly for $\beta < \alpha$).

$X = B_\alpha$, where B is a subset of τ_β.

We define a support S_b for each element b of τ_γ, $\gamma < \alpha$: we select one element from each orbit under γ-allowable permutations, define S_b as the designated support for b, then for each other b' in the orbit choose a γ-allowable permutation π such that $\pi(b) = b'$ and define $S_{b'}$ as $\pi[S_b]$.

We describe the computation of an α-support T_b for $\{b\}_\alpha$ (where b is an element of B) from the β support S_b whose construction is described above This was quite elaborate in the previous version, but here we can simply describe it as $S_b^{\{\alpha\}}$: note that a lot more information can be read from T_b because we can consult f maps with higher index to find additional information about near-litters.

It is given that T_b is computed from S_b, but what we actually need is something like the specification of T_b being computable from the specification of S_b: we cannot in fact compute the specification of T_b from the specification of S_b, but we argue that there are $< \mu$ possibilities for specifications of supports T_b given the specification of S_b.. This is much easier than in the previous version. T_b is precisely parallel in structure to S_b, and the only additional information in T_b^* is the addition of coding functions or pointers to typed atoms where we look at typed-near litters which are flexible in S_b and not flexible in T_b. There will be a small collection of insertions of coding functions taken from collections of size $< \mu$ by inductive hypothesis.

For each $b \in B$ there is a support S_b chosen as above, from which the support T_b can be computed as described above. If $b' \in B$ is in the range of the same coding function χ_{b,S_b} as b, $S_{b'}$ is $\pi[S_b]$ for some β-allowable π with $\pi(b) = b'$. If we have the further condition that T_b and $T_{b'}$ have the same specification, it follows that there is a permutation π_2 such that $\pi_2[T_b] = T_{b'}$. Note that $(\pi_2)_\beta[S_b] = S_{b'}$, from which it follows that $\pi^{-1} \circ (\pi_2)_\beta$ fixes b, since it fixes all elements of S_b, so $b' = \pi(b) = (\pi_2)_\beta(b)$, from which it follows that $\pi_2(\{b\}_\alpha) = \{b'\}_\alpha$ so $\{b\}_\alpha$ and $\{b'\}_\alpha$ are in the range of the same coding function $\chi_{\{b\}_\alpha,T_b}$. Now there are $< \mu$ possible specifications of a coding function χ_{b,S_b} followed by a specification for T_b, so by this procedure we describe a family of $< \mu$ coding functions $\chi_{\{b\}_\alpha,T_b}$ whose range covers all type α singletons of elements of B.

We claim that $\chi_{X,S}$ can be defined in terms of the orbit of S in the allowable permutations and the set of coding functions $\chi_{\{b\}_\alpha,T_b}$. There are $< \mu$ coding functions of this kind, and we have shown above that there are $< \mu$ orbits in the α-strong supports under allowable permutations, so this will imply that there are $\leq \mu$ elements of type α (it is obvious that there are $\geq \mu$ elements of each type). Of course we get $\leq \mu$ codes for each $\beta < \alpha$, but we know that $\lambda < \kappa < \mu$.

The definition that we claim works is that $\chi_{X,S}(U) = B'_\alpha$, where B' is the set of all $\bigcup(\chi_{\{b\}_\alpha,T_b}(U') \cap \pi_\beta)$ for $b \in B$ and U' end extending U. Clearly this definition depends only on the orbit of S and the set of coding functions T_b derived from B as described above. Before we know that this is actually the coding function desired, we will write it as $\chi^*_{X,S}$.

The function we have defined is certainly a coding function, in the sense that $\chi^*_{X,S}(\pi[S]) = \pi(\chi^*_{X,S}(S))$. What requires work is to show that $\chi^*_{X,S}(S) = X$, from which it follows that it is in fact the intended function.

Clearly each $b \in B$ belongs to $\chi^*_{X,S}(S)$ as defined, because $b = \bigcup(\chi_{\{b\}_\alpha,T_b}(T_b) \cap \tau_\beta)$, and T_b end extends S.

An arbitrary $c \in \chi^*_{X,S}(S)$ is of the form $\bigcup(\chi_{\{b\}_\alpha,T_b}(U) \cap \tau_\beta)$, where U end extends S and of course must be in the orbit of T_b under allowable permutations, so some $\pi_0[T_b] = U$. Now observe that $\pi_0[S] = S$, so $\pi_0(X) = X$, so $(\pi_0)_\beta``B = B$. Further $(\pi_0)_\beta(b) = c$, so in fact $c \in B$ which completes the argument. The assertion $(\pi_0)_\beta(b) = c$ might be thought to require verification: the thing to observe is that $c = \bigcup(\chi_{\{b\}_\alpha,T_b}(U) \cap \tau_\beta) = \bigcup(\pi_0(\chi_{\{b\}_\alpha,T_b}(S) \cap \tau_\beta) = \bigcup(\pi_0(\{b\}_\alpha) \cap \tau_\beta) = \bigcup(\{(\pi_0)_\beta(b)\}_\beta \cap \tau_\beta) = (\pi_0)_\beta(b)$

This completes the proof: any element of a type is determined by a support (of which there are μ) and a coding function whose domain includes a strong support (there are $< \mu$ of these, so a type has no more than μ elements (and obviously has at least μ elements).

Note for the formal verification project: I think the latest revisions are closer to the standard needed for the Lean verification project.

4.3 The structure is a model of predicative TTT

There is then a very direct proof that the structure presented is a model of predicative TTT (in which the definition of a set at a particular type may not mention any higher type). Use E for the membership relation $\in_{TTT}$ of the structure defined above. It should be evident that $xEy \leftrightarrow \pi_\beta(x)E\pi(y)$, where x is of type β, y is of type α, and π is an α-allowable permutation.

Suppose that we are considering the existence of $\{x : \phi^s\}$, where ϕ is a formula of the language of TST with $\in$ translated as E, and s is a strictly increasing sequence of types. The truth value of each subformula of ϕ will be preserved if we replace each u of type $s(i)$ with $\pi_{A_{s,i}}(u)$, where $A_{s,i}$ is the set of all s_k for $i \leq k \leq j+1$ [x being of type $s(j)$, and there being no variables of type higher than $s(j+1)$]: $\pi_{A_{s,i}}(x)E\pi_{A_{s,i+1}}(y)$ is equivalent to $(\pi_{A_{s,i+1}})_{s(i)}(x)E\pi_{A_{s,i+1}}(y)$, which is equivalent to xEy by the observation above. The formula ϕ will contain various parameters a_i of types $s(n_i)$ and it is then evident that the set $\{x : \phi^s\}$ will be fixed by any $s(j+1)$-allowable permutation π such that $\pi_{A_{s,n_i}}$ fixes a_i for each i. But this means that $(s(j+1), s(j), \{x : \phi^s\})$ is symmetric and belongs to type $s(j+1)$: we can merge the supports of the a_i's (with suitable raising of indices) into a single $s(j+1)$-support. Notice that we assumed the predicativity condition that no variable more than one type higher than x appears (in the sense of TST).

This procedure will certainly work if the set definition is predicative (all bound variables are of type no higher than that of x, parameters at the type of the set being defined are allowed).

There are easier proofs of the consistency of predicative tangled type theory; there is a reason of course that we have pursued this one.

It should be noted that the construction given here is in a sense a Frankel-Mostowski construction, though we have no real need to reference the usual FM constructions in ZFA here. Constructions analogous to Frankel-Mostowski constructions can be carried out in TST using permutations of type 0; here we are doing something much more complicated involving many permutations of type -1 which intermesh in precisely the right way. Our explanation of our technique is self-contained, but we do acknowledge this intellectual debt.

Note for the formal verification project: We note that in order to avoid metamathematics, we actually suggest proving finitely many instances of comprehension with typed parameters from which the full comprehension scheme can be deduced. That there are such finite schemes (mod the infinite sequence of types) is well-known. For the project, a list should be provided here.

4.4 Impredicativity: verifying the axiom of union

What remains to complete the proof is that typed versions of the axiom of set union hold. That this is sufficient is a fact about predicative type theory. If we have predicative comprehension and union, we note that for any formula ϕ, $\{\iota^k(x) : \phi(x)\}$ will be predicative if k is taken to be large enough, then application of union k times to this set will give $\{x : \phi(x)\}$. $\iota(x)$ here denotes $\{x\}$. It is evidently sufficient to prove that unions of sets of singletons exist.

So what we need to show is that if $\alpha > \beta > \gamma$ and $G \subseteq \tau_\gamma$, and

$$\{\{g\}_\beta : g \in G\}_\alpha$$

is symmetric (has an α-support, so belongs to τ_α), then G_β is symmetric (has a β-support, so belongs to τ_β).

Suppose that $\{\{g\}_\beta : g \in G\}_\alpha$ is symmetric. It then has a strong support S. We claim that $S_{(\beta)}$ (definition of this given in previous subsection, adaptable to supports as sets) is a β-support for G_β.

Any $g \in G$ has a strong γ-support T which extends $(S_{(\beta)})_{(\gamma)}$.

Suppose that the action of the β-allowable permutation π fixes $S_{(\beta)}$.

Our plan is to use freedom of action technology to construct a permutation π^* whose action on S is the identity and whose action on $T^{\{\alpha,\beta\}}$ precisely parallels the action of π on $T^{\{\beta\}}$.

If this is accomplished, then the action of π^* fixes S and so fixes

$$\{\{g\}_\beta : g \in G\}_\alpha,$$

while at the same time $(\pi^*_\beta)_\gamma$ agrees with π_γ on G. This implies that $\pi_\gamma(g) \in G$ (and the same argument applies to π^{-1}) so π fixes $\{\{g\}_\beta : g \in G\}$.

Close up the γ-support T under the processes of action of π and inclusion of atoms at which π acts exceptionally to obtain T^*.

We construct the allowable permutation π^* by Freedom of Action so that the action of $(\pi^*_\beta)_\gamma$ on atomic and flexible items in T^* agrees with the action of π_γ on T^* and the action of π_* fixes atomic and flexible items in S. On any non-flexible litter L in S, π^*_β acts correctly because it acts correctly on a support of the inverse image of L under the appropriate f map (fixing all of its elements). The tricky case seems to require a little extra attention to the action on T^*: if a non-flexible litter has inverse image u under $f^{-1,\gamma}$, it is mapped by π to something with inverse image v under $f^{-1,\gamma}$, we arrange for the approximation generating π^* to induce π^*_β to map $\{u\}_\beta$ to $\{v\}_\beta$. Thus $(\pi^*_\beta)_\gamma$ maps g to $\pi_\gamma(g)$ as required for the argument above. That said, any non-flexible item is sent to its image under the appropriate derivative of π because a support is acted on correctly and there will be no exceptional actions of derivatives of π^* disagreeing with exceptional actions of π because T^* is closed under exceptional actions of π in litters. This completes the argument.

NOTE: Difficult interactions with S are avoided because an incompatibility of π with fixing S would involve moving most elements of a litter in the range of $f_{-1,\beta}$, and while π may do this, nothing in the definition of π^* can force this to happen; there is no conflict between the conditions imposed by S and the conditions imposed by T^*.

Note for formal verification project: This is converging to a full description at the level needed for formalization...

5 Conclusions, extended results, and questions

[I have copied in the conclusions section of an older version, but what it says should be about right, and may require some revisions to fit in this paper. I also added the bibliography, which again is probably approximately the right one.]

This is a rather boring resolution of the NF consistency problem.

NF has no locally interesting combinatorial consequences. Any stratified fact about sets of a bounded standard size which holds in ZFC will continue to hold in models constructed using this strategy with the parameter κ chosen large enough. That the continuum can be well-ordered or that the axiom of dependent choices can hold, for example, can readily be arranged. Any theorem about familiar objects such as real numbers which holds in ZFC can be relied upon to hold in our models (even if it requires Choice to prove), and any situation which is possible for familiar objects is possible in models of *NF*: for example, the Continuum Hypothesis can be true or false. It cannot be expected that *NF* proves any strictly local stratified result about familiar mathematical objects which is not also a theorem of ZFC.

Questions of consistency with NF of global choice-like statements such as "the universe is linearly ordered" cannot be resolved by the method used here (at least, not without major changes). One statement which seems to be about big sets can be seen to hold in our models: the power set of any well-orderable set is well-orderable, and more generally, beth numbers are alephs. We indicate the proofs: a relation which one of our models of TTT thinks is a well-ordering actually is a well-ordering, because the models are countably complete; so a well-ordering with a certain support has all elements of its domain sets with the same support (a permutation whose action fixes a well-ordering has action fixing all elements of its domain), and all subsets of and relations on the domain are sets with the same support (adjusted for type differential), and this applies further to the well-ordering of the subsets of the domain which we find in the metatheory. Applying the same result to sets with well-founded extensional relations on them proves the more general result about beth numbers. This form of choice seems to allow us to use choice freely on any structure one is likely to talk about in the usual set theory. It also proves, for example, that the power set of the set of ordinals (a big set!) is well-ordered.

NF with strong axioms such as the Axiom of Counting (introduced by Rosser in [13], an admirable textbook based on *NF*), the Axiom of Cantorian Sets (introduced in [4]) or my axioms of Small Ordinals and Large Ordinals (introduced in my [6] which pretends to be a set theory textbook based on *NFU*) can be obtained by choosing λ large enough to have strong partition properties, more or less exactly as I report in my paper [7] on strong axioms of infinity in NFU: the results in that paper are not all mine, and I owe a good deal to Solovay in that connection (unpublished conversations and [17]).

That NF has α-models for each standard ordinal α should follow by the same methods Jensen used for NFU in his original paper [10]. No model of NF can contain all countable subsets of its domain; all well-typed combinatorial consequences of closure of a model of TST under taking subsets of size $< \kappa$ will hold in our models, but the application of compactness which gets us from TST + Ambiguity to NF forces the existence of externally countable proper classes, a result which has long been known and which also holds in NFU.

We mention some esoteric problems which our approach solves. The Theory of Negative Types of Hao Wang (TST with all integers as types, proposed in [21]) has ω-models; an ω-model of NF gives an ω-model of TST immediately. This question was open.

In ordinary set theory, the Specker tree of a cardinal is the tree in which the top is the given cardinal, the children of the top node are the preimages of the top under the map $(\kappa \mapsto 2^\kappa)$, and the part of the tree below each child is the Specker tree of the child. Forster proved using a result of Sierpinski that the Specker tree of a cardinal must be well-founded (a result which applies in ordinary set theory or in NF(U), with some finesse in the definition of the exponential map in NF(U)). Given Choice, there is a finite bound on the lengths of the branches in any given Specker tree. Of course by the Sierpinski result a Specker tree can be assigned an ordinal rank. The question which was open was whether existence of a Specker tree of infinite rank is consistent. It is known that in NF with the Axiom of Counting the Specker tree of the cardinality of the universe is of infinite rank. Our results in this paper can be used to show that Specker trees of infinite rank are consistent in bounded Zermelo set theory with atoms or without foundation (this takes a little work, using the way that internal type representations unfold in TTT and a natural interpretation of bounded Zermelo set theory in TST; a tangled web as described above would have range part of a Specker tree of infinite rank). A bit more work definitely gets this result in ZFA, and we are confident that our permutation methods can be adapted to ZFC using forcing in standard ways to show that Specker trees of infinite rank can exist in ZF.

We believe that NF is no stronger than TST + Infinity, which is of the same strength as Zermelo set theory with separation restricted to bounded formulas. Our work here does not show this, as we need enough Replacement for existence of $\beth_{\omega_1}$ at least. We leave it as an interesting further task, possibly for others, to tighten things up and show the minimal strength that we expect holds.

Another question of a very general and amorphous nature which remains is: what do models of NF look like in general? Are all models of NF in some way like the ones we describe, or are there models of quite a different character? There are very special assumptions which we made by fiat in building our model of TTT which do not seem at all inevitable in general models of this theory.

I am not sure that all references given here will be used in this version.

References

[1] Crabbé, Marcel, reference for SF interpreting NFU.

[2] Forster, T.E. [1995] Set Theory with a Universal Set, exploring an untyped Universe Second edition. Oxford Logic Guides, Oxford University Press, Clarendon Press, Oxford.

[3] Hailperin, finite axiomatization

[4] Henson, C.W. [1973a] Type-raising operations in NF. Journal of Symbolic Logic 38 , pp. 59-68.

[5] Holmes, M.R. "The equivalence of NF-style set theories with "tangled" type theories; the construction of omega-models of predicative NF (and more)". *Journal of Symbolic Logic* 60 (1995), pp. 178-189.

[6] Holmes, M. R. [1998] Elementary set theory with a universal set. volume 10 of the Cahiers du Centre de logique, Academia, Louvain-la-Neuve (Belgium), 241 pages, ISBN 2-87209-488-1. See here for an on-line errata slip. By permission of the publishers, a corrected text is published online; an official second edition will appear online eventually.

[7] Holmes, M. R. [2001] Strong Axioms of infinity in NFU. Journal of Symbolic Logic, 66, no. 1, pp. 87-116.
("Errata in 'Strong Axioms of Infinity in NFU' ", JSL, vol. 66, no. 4 (December 2001), p. 1974, reports some errata and provides corrections).

[8] Kemeny thesis on strength of TST

[9] Jech, Thomas, *Set theory*, Academic Press 1978, pp. 199-201.

[10] Jensen, R.B. "On the consistency of a slight(?) modification of Quine's NF". *Synthese* 19 (1969), pp. 250-263.

[11] Quine on ordered pairs

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

[13] Rosser, J. B. [1978] Logic for mathematicians, second edition. Chelsea Publishing.

[14] Russell, Principles of Mathematics

[15] Russell and Whitehead, Principia Mathematica

[16] Scott, Dana, "Definitions by abstraction in axiomatic set theory", *Bull. Amer. Math. Soc.*, vol. 61, p. 442, 1955.

[17] Solovay, R, "The consistency strength of NFUB", preprint on `arXiv.org`, `arXiv:math/9707207 [math.LO]`

[18] Specker, E.P. "The axiom of choice in Quine's new foundations for mathematical logic". *Proceedings of the National Academy of Sciences of the USA* 39 (1953), pp. 972-975.

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

[20] Tarski, first description of TST

[21] Wang, H. [1952] Negative types.

[22] Wiener, Norbert, paper on Wiener pair