New Foundations is consistent

M. Randall Holmes and Sky Wilshaw

January 8, 2025

Contents

1 Introduction

I (Sky) am going to write annotations in this format. I will write suggestions for additions in red text.

Holmes: I'll write remarks in magenta.

1.1 Introductory note

We are presenting an argument for the consistency of Quine's set theory New Foundations (NF) (see [12]). The consistency of this theory relative to the usual systems of set theory has been an open question since the theory was proposed in 1937, with added urgency after Specker showed in 1954 that NF disproves the Axiom of Choice (see [19]). Jensen showed in 1969 that NFU (New Foundations with extensionality weakened to allow urelements) is consistent and in fact consistent with Choice (and also with Infinity and with further strong axioms of infinity) (see [10]).

The first author showed the equiconsistency of NF with the quite bizarre system TTT (Tangled Type Theory) in 1995 [see [6], but this paper is not really recommended; the presentation of its results here is better], which gave a possible approach to a consistency proof. Since 2010, the first author has been attempting to write out proofs, first of the existence of a "tangled web of cardinals" in Mac Lane set theory [not defined in this paper] and then directly of the existence of a model of tangled type theory. These proofs are difficult to read, insanely involved, and involve the sort of elaborate bookkeeping which makes it easy to introduce errors every time a new draft is prepared. The second author (with the assistance of others initially) has formally verified the proof of the first author of the existence of a model of TTT (see [24]), and so of the consistency of New Foundations, in the Lean proof verification system, and in the process has suggested minor corrections and considerable formal improvements to the argument originally proposed by the first author. The second author reports that the formalized proof is still difficult to read and insanely involved with nasty bookkeeping. Both authors feel that there ought to be a simpler approach, but the existing argument at least strongly resists attempts at simplification.

All the theories mentioned here are discussed (and referenced) in the next section.

Any remarks in the first person singular may be attributed to the first author.

1.2 Acknowledgements

The first author thanks Robert Solovay, who read a number of early versions of a related argument for Con(NF) and offered just criticisms which I have tried to take to heart. I also thank Thomas Forster and Asaf Karagila, who have endured attempts of mine to present various relatives of this argument at length. Further, I thank the members of the group of students at Cambridge

who attempted formalization of an earlier version of this proof in Summer 2022. We would like to thank Peter Lumsdaine for his work with the students in the summer project. Cambridge students who participated in the summer project (including the second author) were funded partly by DPMMS, and Queens' College, but principally by the Wes and Margaret foundation, who had also supported work on New Foundations in the past, and we are grateful.

1.3 Dated remarks on versions

initial note for this series: The formal verification of the results in section 3 and 4 is complete. Sky's Lean formalization concludes with a proof that there is a structure for the language of TTT which satisfies all typed versions of the statements in the Hailperin axiomatization of NF. Such a structure is a model of TTT, and the existence of a model of TTT implies the consistency of New Foundations. Thus we know that the construction in the paper is correct in general terms and we need to rewrite it for publication to parallel the formalized proof as much as possible: we certainly know that the conclusion of the paper expressed in the title is true.

initial note for the previous series: 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 in our view (at least in the view of the first author) avoid metamathematics as far as possible, 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. **This program has been completed, on the Lean side.**

1/8/2025: reviewed and slightly revised the proof of the existence of position functions. Further input from Sky at the meeting.

1/1/2025: Sky proposed changes to the beginning of section 3, which are incorporated.

12/31/2024: Pulled out the section natural models and theories TST_n, because no use seems to be made of those concepts in the paper as it stands.

12/17/2024: Some corrections in sections 2.4.2 and 2.4.3. I am contemplating deleting these sections, depending on how revisions of the conclusions and questions section go.

I articulated section 5 into subsections, and removed the need to refer to sections 2.4.2 and 2.4.3 of the original version, so they are now commented out (though available to be reinserted if desired). This means that the notion of tangled web of cardinals is not defined nor discussed in this paper, though it is alluded to briefly in the introductory remarks.

I need some references to results which I use in the new brief argument for existence of Specker trees of infinite rank in models of bounded Zermelo set theory.

Sky has indicated that there may be fencepost errors to be corrected in my discussion of the Hailperin axioms; I believe I have resolved (or in one case finessed) these.

10/16/2024: Attempting to add a version of the proof that there are α-models which is needed to support comments in the results and conclusions section. It is not a terribly accessible result, and it might not be easy to see that Jensen's proof applies in our context. I note here that while I believe the proof I give for α-models is correct, it does need to be written more elegantly, and I think the argument I give is a blunt instrument: I think the cofinality of λ which is needed may be just $|\alpha|$ and I'll consult my sources to see if I can add finesse to achieve that, if possible.

Note added later on the 16th: It appears that Jensen's original argument is basically the same as mine, so there is no expectation that there is a way to reduce the required cofinality of λ globally, or at least we are under no pressure to supply one. But it appears that Marcel did show the existence of ω-models of NFU using $\beth_{\omega_1}$ instead of $\beth_{c+}$, which is what my argument (or Jensen's) appears to require.

10/17/2024: Further improvements in the language of the subsection on α-models. The reason this section is needed, though the proof given differs only superficially from Jensen's in a certain sense, is that it needs to be made really clear that this method adapts to our framework for constructing models of NF. I needed to write it out to be sure of this, so I suspect a reader should need this.

10/8/2024: This is exactly the same text as the version submitted to arxiv, unless I made some error in cutting out comments, except that arxiv insisted that I remove the colored remarks. I have deleted lots of old comments from the source. I have also set things up so that the remaining colored comments can be turned on or off: in the web page version, they are turned on.

9/30/2024: This version will be posted to arxiv today. Changes from 9/10 are minor. A lot of the version notes have been edited out, and all the editorial remarks, as arxiv requested.

Mod the possibility of further remarks from Sky, this is the new version for Arxiv. Arxiv readers should note that the proof of Freedom of Action is completely rewritten by Sky Wilshaw in this version. It is in some sense the same argument as my original argument, or close to being so, but it is much easier to read. The last of the series of drafts before Sky's revision of the FoA proof is preserved on my web page for the time being.

9/3/2024: Starting new version for Arxiv with editing comments suppressed, reflecting the rewrite of the proof of Freedom of Action by the second author and various minor corrections and rearrangements.

8/16/2024: New revisions by Sky and Holmes responses. Extended type indices now always have minimum -1.

8/13/2024: Working copy for the meeting with Sky on this date.

8/11/2024: This version incorporates a substantial bookkeeping change. The second components of support conditions now have -1 as minimal element, which improves uniformity in some boundary cases, but which necessitates a lot of changes to indexing and notation here and there.

7/26/2024: This contains Sky's complete rewrite of the proof of Freedom of Action, without changes to subsequent text which it entails. It also includes redefinition of the position of function to act on singletons of atoms instead of atoms, to avoid irritating type conflicts between the definition of the position function and the definition of support conditions.

7/2/2024: Starting a revision effort for the month of July, to support reconciliation of the paper with the formal proof with an eye to submission for publlcation.

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[1] (*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 $\mathtt{type}(x)$: this will be a nonnegative integer. A countably infinite supply of variables of each type is supposed. We provide a bijection $(x \mapsto x^+)$ from variables to variables of positive type satisfying $\mathtt{type}(x^+) = \mathtt{type}(x) + 1$.[2]

An atomic equality sentence $x = y$ is well-formed iff $\mathtt{type}(x) = \mathtt{type}(y)$. An atomic membership sentence $x \in y$ is well-formed iff $\mathtt{type}(x) + 1 = \mathtt{type}(y)$.

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, complex terms 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 necessarily uniquely defined when ϕ is uniformly false by defining $\{x^i : \phi\}$ as $\emptyset^{i+1}$ in this case. The superscript is sometimes omitted from $\emptyset^i$ (and other constants parametrized by types) when the type can be deduced from context.

[1] led by Maurice Boffa, to whom the first author owes a profound debt, for being harbored in his group at the beginning of my career, which made it possible for me to launch my research program without distractions attendant on looking for a job.

[2] We do *not* furnish our variables with type superscripts. One could follow the convention that if a variable has a natural number superscript, this determines its type, and that the effect of $^+$ on a superscripted variable is to increment the superscript, but we do not expect variables to have superscripts.

2.1.1 Typical ambiguity

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

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 ([16]), 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* ([15]). 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 ([23]), 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 the 1930s[3].

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.

[3]We regard it as an important feature of TST that the nature of type 0 is left completely undetermined, so we do not regard descriptions of arithmetic of order ω appearing about the same time as part of this history. Arithmetic of order ω is a similar typed theory but lacks the ambiguity of interest here.

It was shown around 1950 (in [9]) 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

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 equinumerosity 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

$|\iota\text{“}X| = |\mathcal{P}(A)|$; 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 $|\iota\text{“}X| = |\mathcal{P}(V)|$, because $|\iota\text{“}V| < |\mathcal{P}(V)| \leq |V|$, and of course there is no set larger than V in its type.

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.[4]

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 variable 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. We do this in a way which is usual now but not necessarily identical to what is done in the original paper. 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 $\sigma(\text{`}x\text{'}) = \sigma(\text{`}y\text{'})$ and for each atomic subformula '$x \in y$' of ϕ we have $\sigma(\text{`}x\text{'}) + 1 = \sigma(\text{`}y\text{'})$. 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 ([4]), refined by others, and discussed in some detail in a subsequent subsection of this section, because it is relevant to the Lean formalization. One obtains a finite axiomatization of NF by analogy

[4]There is a persistent rumor that Quine's set theory in its original version was inconsistent. This is not the case (and in fact we show here that the system of the 1937 paper in which the system was introduced is consistent). The truth behind the rumor is that the system of the first edition of Quine's book *Mathematical Logic* ([13]), in which proper classes were added to NF, was inconsistent; the inconsistency was corrected in the second edition of 1951, and the results of this paper show that the system of the 1951 book is consistent. We do not discuss the systems with proper classes further here.

with the method of finitely axiomatizing von Neumann-Gödel-Bernays predicative class theory. It should further be noted that the first thing one does with any 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. It is also worth noting that the collection of all typed instances of the Hailperin axioms is an axiomatization of TST, not a finite axiomatization, but an axiomatization with finitely many templates.

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 ([20], 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 prove a version of Specker's result using concepts from this paper below.

In [19], 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 $|\iota\text{“}V| < |\mathcal{P}(V)| \leq |V|$ (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 [6], 1995, the first 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 $\mathtt{type}(\text{`}x\text{'}) < \lambda$; 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 $\{s_i\}_{\mathbb{N}}$ 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 $\mathtt{type}(\text{`}x\text{'}) = \mathtt{type}(\text{`}y\text{'})$ and whose atomic sentences '$x \in y$' satisfy $\mathtt{type}(\text{`}x\text{'}) < \mathtt{type}(\text{`}y\text{'})$, 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 $(\forall x^\alpha : x^\alpha \notin \emptyset^{\alpha,\beta})$ 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 2.1 (Holmes)**.** TTT(U) is consistent iff NF(U) is consistent.

Proof.

1. 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.

2. (a) 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). Our first aim is to construct a model of TST(U) in which $\phi \leftrightarrow \phi^+$ holds for each sentence $\phi \in \Sigma$.

   (b) Let n be the smallest natural number such that all variables which occur in Σ have type less than n. 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 $\mathtt{rng}(s \restriction \{0,\ldots,n-1\}) = A$: there are no more than $2^{|\Sigma|}$ compartments. By Ramsey's theorem, there is an infinite set $H \subseteq \lambda$ homogeneous for this partition which includes the range of a strictly increasing sequence h.

(c) Any strictly increasing sequence s determines a model of TST(U) in which each type i is implemented as type $s(i)$ of our fixed model of TTT(U) and membership in the derived model of TST(U) of type i objects in type $i+1$ objects is implemented as membership in the fixed model of TTT(U) of type $s(i)$ objects in type $s(i+1)$ objects.

(d) This model determined in this way by h satisfies $\phi \leftrightarrow \phi^+$ for each $\phi \in \Sigma$. This achieves our first aim.

(e) But this implies by compactness that the full Ambiguity Scheme $\phi \leftrightarrow \phi^+$ is consistent with TST(U). We could at this point say that NF(U) is consistent by the 1962 result of Specker ([20]), but we will continue and indicate how to prove this directly.

(f) 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(U), by constructing a term model using the Hilbert symbol in the natural way, then identifying all terms with their type-raised versions (an element of the term model is a class of terms $(\epsilon x : \phi)$ in the language of TST(U) which is obtained from a single such term by all possible upward or downward shifts of type by a constant amount: we write this $[(\epsilon x : \phi)]$).

(g) All statements in the resulting type-free theory can be decided: the truth value of a closed atomic formula $[(\epsilon x : \phi)]\, R\, [(\epsilon y : \psi)]$ in the model of NF(U) is determined by replacing each side of the formula with one of its elements in such a way as to get a well-typed formula, which will always be possible if types are taken high enough, and reading the truth value of the resulting formula from the model of TST(U) with Ambiguity and Hilbert symbol; R is either $=$ or $\in$ [note that the relation interpreting equality in the term model is a nontrival equivalence relation]. Truth values of more complex statements are determined in standard ways. That this makes axioms of NF(U) true should be clear.

(h) 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 in paragraphs (a)–(e) adapts to show consistency of TST(U) plus Ambiguity with the Hilbert symbol.

(i) We remark that our method of proving (a version of) the Specker ambiguity result is distinctly different from the original method. Strictly

speaking, we have only proved Specker's result for TST(U) + Ambiguity + existence of a Hilbert symbol, but that is all we need for our result.

(j) From this it follows that NF(U) is consistent.

□

Theorem 2.2 (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 There are α-models for every α

This section is devoted to adapting a result of Jensen to our context. Jensen proved in [10] that for any infinite ordinal α of the metatheory, there is a model of NFU + Infinity in which there is a well-ordering of type α. This is of particular interest in the case $\alpha = \omega$, which establishes that the natural numbers of a model can correspond precisely to the natural numbers of the metatheory.

Nothing in our construction of a model of TTT below depends on this section: the fact that an existence of an ω-model of NF follows from our main result is significant for extensions of our results and uses this theorem.

A relation R of one of the theories under discussion is a set of ordered pairs of the theory in question; note that it corresponds to a set of ordered pairs of the metatheory with the same field. If this relation corresponding to R is a well-ordering, we can talk about its order type both in terms of the theory under consideration and in terms of the metatheory, though ordinals of the theory under consideration and the metatheory may not be the same.

In any of the theories under discussion, an ordinal is taken to be an equivalence class of well-orderings under similarity. The relations which belong to an ordinal of one of these theories may fail to be well-orderings in terms of the metatheory. If α is an ordinal of the metatheory, we say that α is implemented in a type of one of the theories under discussion exactly if there is an ordinal in that type (in the sense of the theory) whose elements in fact correspond to well-orderings of order type α in the sense of the metatheory, in the sense outlined in the previous paragraph. It is straightforward to show that if α is implemented in a type, the ordinal implementing it will contain all well-orderings in the sense of the theory of the appropriate lower type whose order type is α in the sense of the metatheory.

Theorem 2.3. For any infinite ordinal α of the metatheory, if there is a model of TTT(U) in which the ordinal α has an implementation α^s for each strictly increasing sequence s of types, by which we mean an ordinal in type s_4 whose elements of type s_3 correspond to well-orderings of order type α of subsets of type s_0, using types s_1 and s_2 for the intermediate pairing machinery, and which is such that the ordinal λ indexing the types is a strong limit cardinal of cofinality $> |\alpha|^{|\alpha|}$, then there is a model of NF(U) in which the ordinal α is implemented, which further satisfies [after neglect of types] any sentence in the language of TST(U) whose translation into terms of each increasing sequence of types in the model of TTT(U) is satisfied in the model of TTT(U).

Proof. The proof is a modified version of the proof given above for consistency of NF(U), using a more powerful partition principle. We do not claim originality, referring the reader to [10], but our presentation is not the same, and it is adapted to talk of tangled type theory.

We extend the language of TTT(U) to include constant terms α^s and β^s for each ordinal $\beta < \alpha$ of the metatheory for each sequence s as described above. We similarly extend the language of TST(U) (in this case only needing constants α^i and β^i for $\beta < \alpha$ for each $i \geq 4$), and note that the translation of a

sentence of TST(U) with language augmented with these ordinal constants into terms of a strictly increasing sequence of types in TTT(U) is straightforward, and we allow use of these constants in the comprehension axiom of TTT(U) as usual. We further include the Hilbert symbol in our language, with the same restriction stated above that the Hilbert symbol may not appear in instances of comprehension, noting our intention to construct a term model.

We consider partitions of $[\lambda]^n$ determined not by a finite set Σ of sentences of the language of TST(U) using types $< n$ with compartments determined by truth values assigned to the elements of Σ, but determined by the set Σ_n of *all* Hilbert symbols of the form $(\epsilon\beta^i : \beta^i < \alpha^i \wedge \phi)$ in the language of TST(U) augmented with Hilbert symbols and constants for the ordinals $\leq \alpha$ in each type which contains ordinals, using only types less than n, with the compartment in the partition containing $A \in [\lambda]^n$ determined by the ordinal values $< \alpha$ assigned to each of these terms in our model of TTT(U) when types $0-(n-1)$ are replaced by the types in A in increasing order in each of the Hilbert symbols in Σ_n.

There are no more than $|\alpha|$ such Hilbert symbols, and no more than $|\alpha|^{|\alpha|}$ compartments in each of these partitions.

We claim that we can construct a complete assignment of values to such Hilbert symbols in the extended TST(U) language (which will induce a complete assignment of truth values to sentences of the extended language:

$$(\epsilon\beta^i : \beta^i < \alpha \wedge \beta^i < 2 \wedge (\beta = 1 \leftrightarrow \phi))$$

codes a sentence ϕ into Σ_n if it mentions only types below n, because the assignment of ordinal value to this term corresponds to the assignment of truth value to ϕ) which is ambiguous [invariant under displacement of types] and consistent with any statement holding at all sequences of types in the model of TTT(U) (so for example we will get extensionality if the model of TTTU is actually a model of TTT), and in fact for any fixed n all these assignments for Hilbert symbols mentioning only types less than n will be realized at some sequence of types in the model of TTT(U) [verifying that this assignment of values is consistent and in effect a complete description of a model of NF(U)].

Further, it will be clear that the model of NFU will contain no ordinal less than what it calls α other than the ordinals which it calls β for $\beta < \alpha$ in the metatheory, and the natural well-ordering on these ordinals is realized in the model of NF(U) and corresponds to a well-ordering of order type α in the metatheory. The reason for this is that the model of NFU is a term model, and any term denoting an ordinal less than α can be slightly modified to the exact form used in defining the partitions, and the process will assign that term a concrete value less than α, an actual ordinal $\beta < \alpha$ of the metatheory, so the order on the ordinals $< \alpha$ in the model of NF(U) has order type α in the metatheory.

The Erdös-Rado partition theorem tells us that for any fixed n, there are arbitrarily large homogeneous sets for our partition of $[\lambda]^n$ determined by values of Hilbert symbols in Σ_n. For any cardinal $\mu < \lambda$, the Erdös-Rado theorem tells us directly that the partition of $\exp^n(\mu) < \lambda$ defined in the way indicated above

has a homogeneous set of size μ^+, and λ is a limit cardinal, so the choice of μ is unbounded below λ.

Further, because the cofinality of λ is greater than $|\alpha|^{|\alpha|}$, there is a specific assignment of values to the Hilbert symbols in Σ_n such that the upper bound on the size of homogenous sets for the associated partition which realize this assignment of values is λ (briefly, we say that this assignment is associated with arbitrarily large homogeneous sets). We can then pass to Σ_{n+1} and construct homogeneous sets of arbitrarily large size below λ for the partition of Σ_{n+1} by constructing them inside homogenous sets of sufficiently large size for the partition determined by Σ_n associated with this specific assignment of values to the terms in Σ_n, and each of these homogeneous sets of course realizes an assignment of values to the terms in Σ_{n+1} extending the given one to terms in Σ_n. We can then choose a specific assignment of values to the terms in Σ_{n+1} extending the assignment of values to terms in Σ_n which is associated with homogeneous sets of arbitrarily large size, due to cofinality considerations, and this can be repeated through ω steps to produce a complete assignment of values to the Hilbert symbols under consideration, with the properties outlined above. □

2.5 An axiomatization of TST with finitely many templates

This entire subsection needs to be checked carefully. The calculations here are nasty (Holmes) This note remains in the "clean" version because the task remains. I note for readers that nothing in the rest of the paper hinges on details here: all that is used is the bare fact that Hailperin's axioms are sufficient, which has long been known.

We discuss a finite axiomization of NF derived from that of Hailperin (it is taken from an implementation of Hailperin's axiom set of [4] in Metamath ([2]), and there are minor changes from the original formulation), making the important observation that it actually provides us with an axiomatization of TST with finitely many axiom templates (in the sense that each axiom is a type shifted version of one of a finite set of axioms). Notation introduced in this section is not used in the rest of the paper, and nothing in subsequent sections except a brief remark in the last paragraphs of section 4 depends on anything here. This finite axiomatization is however used in the Lean formalization.

The finite axiomatization of NF takes this form (the definitions inserted are ours, and we have modified the order of the axioms to make the definitions work sensibly). We also present this as an axiomatization of TST with finitely many templates, with the proviso that each typed form of each axiom is asserted:

extensionality axiom: $(\forall x : (\forall y : (\forall z : z \in x \leftrightarrow z \in y) \rightarrow x = y))$

anti-intersection axiom: $(\forall xy : (\exists z : (\forall w : w \in z \leftrightarrow \neg(w \in x \wedge w \in y))))$

singleton axiom: $(\forall x : (\exists y : (\forall z : z \in y \leftrightarrow z = x)))$

definition: $\{x\}$ denotes for each x the set whose only element is x, whose existence is provided by the previous axiom. We define $\iota(x)$ as $\{x\}$ and define $\iota^1(x)$ as $\iota(x)$ and $\iota^{n+1}(x)$ as $\{\iota^n(x)\}$, for each concrete natural number n.

cardinal one axiom: $(\exists x : (\forall y : y \in x \leftrightarrow (\exists z : (\forall w : w \in y \leftrightarrow w = z))))$

definition: We define 1 as the set of all singletons, provided by the previous axiom.

definition: $x|y$ denotes the set z whose existence is provided by the anti-intersection axiom: $z \in x|y \leftrightarrow \neg(z \in x \wedge z \in y)$. We define x^c as $x|x$. We define $x \cap y$ as $(x|y)^c$. We define $x \cup y$ as $x^c|y^c$. We define V as $1|1^c$ (noting that it is straightforward to prove $x|x^c = V$ for any x, since this is the universal set). We define $\{x, y\}$ as $\{x\} \cup \{y\}$. We define (x, y) as $\{\{x\}, \{x, y\}\}$. We define (x, y, z) as $(\{\{x\}\}, (y, z))$. More generally, we define $(x_1, \ldots, x_n)$ as $(\iota^{2n-4}(x_1), (x_2, \ldots, x_n))$ for $n > 2$ a concrete natural number. [The treatment of n-tuples is what makes this axiomatization singularly awful].

cross product axiom: $(\forall x : (\exists y : (\forall z : z \in y \leftrightarrow (\exists wt : z = (w, t) \wedge t \in x))))$

definition: We define $V \times x$ as the set introduced by the previous axiom: $z \in V \times x \leftrightarrow (\exists wt : z = (w,t) \wedge t \in x)$. Note that $V \times V$ is the set of all ordered pairs.

converse axiom: $(\forall x : (\exists y : (\forall zw : (z,w) \in y \leftrightarrow (w,z) \in x)))$

definition: For any set R, we define R^{-1} as the intersection of $V \times V$ with a set introduced by the previous axiom:

$$(\forall zw : (z,w) \in R^{-1} \leftrightarrow (w,z) \in R) \wedge (\forall u : u \in R^{-1} \leftrightarrow (\exists zw : (z,w) = u)).$$

definition: We define $x \times V$ as $(V \times x)^{-1}$ and $x \times y$ as $(x \times V) \cap (V \times y)$.

definitions: We define $\iota^2\text{“}x$ as $(x \times x) \cap 1$. This is the image of x under the double singleton operation.

Note that $\iota^2\text{“}V$ is the equality relation. Define $\iota^{2(n+1)}\text{“}x$ as $\iota^2\text{“}(\iota^{2n}\text{“}x)$.

We define $x_1 \times x_2 \ldots \times x_n$ as $\iota^{2(n-2)}\text{“}x \times (x_2 \times \ldots \times x_n)$.

We define V^1 as V and V^{n+1} as $\iota^{2(n-2)}\text{“}V \times V^n$, for each concrete n. V^n is the set of all n-tuples.

singleton image axiom: $(\forall x : (\exists y : (\forall zw : (\{z\},\{w\}) \in y \leftrightarrow (z,w) \in x)))$.

definition: We define R^ι for any set R as the intersection of a set provided by the previous axiom with 1×1. R^ι is the set whose members are exactly the ordered pairs $(\{z\},\{w\})$ such that $(z,w) \in R$. Let R^{ι^1} be defined as R^ι and $R^{\iota^{n+1}}$ be defined as $(R^{\iota^n})^\iota$.

insertion two axiom: $(\forall x : (\exists y : (\forall zwt : (z,w,t) \in y \leftrightarrow (z,t) \in x)))$

We define $I_2(R)$ as the intersection of a set provided by the previous axiom with $V \times V \times V$.

insertion three axiom: $(\forall x : (\exists y : (\forall zwt : (z,w,t) \in y \leftrightarrow (z,w) \in x)))$

We define $I_3(R)$ as the intersection of a set provided by the previous axiom with $V \times V \times V$.

definition: It seems natural to define $I_1(R)$ as $\iota^2\text{“}V \times R$, but this requires no new axiom.

definition: Define $I_{1,n}(R)$ as $\iota^{2(n-1)}\text{“}V \times R$: this is correct for prepending all possible initial projections to an n-tuple. Then define $I^1_{1,n}(R)$ as $I_{1,n}(R)$ and define $I^{m+1}_{1,n}(R)$ as $I_{1,n+m}(I^m_{1,n}(R))$: this takes into account the fact that the tuples get longer. The intention here is that $I^m_{1,n}$ represents the result of applying the insertion one operation to an n-ary relation m times, yielding an $(n+m)$-ary relation.

Define $I_{2,n}(R)$ and $I^1_{2,n}(R)$ as $I_2(R) \cap V^{n+1}$ [one effect of the intersection with V^{n+1} is to restrict second projections of tuples in R to appropriately iterated singletons], and define $I^{m+1}_{2,n}(R)$ as $I_{2,n+m}(I^m_{2,n}(R))$: this takes

into account the fact that the tuples get longer. The intention here is that $I^m_{2,n}$ represents the result of applying the insertion two operation to an n-ary relation m times, yielding an $(n+m)$-ary relation.

type lowering axiom: $(\forall x : (\exists y : (\forall z : z \in y \leftrightarrow (\forall w : (w, \{z\}) \in x))))$

definition: We define $\mathtt{TL}(x)$ by $(\forall z : z \in \mathtt{TL}(x) \leftrightarrow (\forall w : (w, \{z\}) \in x))$. This is a very strange operation!

We define $\iota^{-1}\text{“}x$ as $\mathtt{TL}(V \times x)$. This is the set of all elements of singletons belonging to x. We can then define $\iota\text{“}x$, the elementwise image of x under the singleton operation, as $\iota^{-1}\text{“}(\iota^2\text{“}(x))$.

Further, we define $\iota^{-(n+1)}\text{“}(x)$ as $\iota^{-1}\text{“}(\iota^{-n}\text{“}(x))$ for each concrete natural number n, and $\iota^{n+1}\text{“}(x)$ as $\iota\text{“}(\iota^n\text{“}(x))$

We develop an important operation step by step.

$\mathtt{TL}(R) = \{z : (\forall w : (w, \{z\}) \in R)\}$

Dually, $(\mathtt{TL}(R^c))^c = \{z : (\exists w : (w, \{z\}) \in R)\}$

Now $(\mathtt{TL}((R^\iota)^c))^c = \{z : (\exists w : (w, \{z\}) \in R^\iota)\}$,

which is the same as $(\mathtt{TL}((R^\iota)^c))^c = \{z : (\exists w : (\{w\}, \{z\}) \in R^\iota)\}$ because all elements of the domain of R^ι are singletons,

which is the same as $(\mathtt{TL}((R^\iota)^c))^c = \{z : (\exists w : (w, z) \in R)\}$

so we define $\mathtt{rng}(R)$ as $(\mathtt{TL}((R^\iota)^c))^c$, and define $\mathtt{dom}(R)$ as $\mathtt{rng}(R^{-1})$.

subset axiom: $(\exists x : (\forall yz : (y, z) \in x \leftrightarrow (\forall w : w \in y \to w \in z)))$

We define $[\subseteq]$ as the intersection of a set provided by the previous axiom with $V \times V$: $[\subseteq]$ is the set of all ordered pairs (x, y) such that $x \subseteq y$.

We define $[\in]$ as $[\subseteq] \cap (1 \times V)$.

This is not our favorite finite axiomatization of NF (or our favorite finite template axiomatization of TST) but it is the one verified in the Lean formalization and also basically the oldest one, so we present a verification of it in outline at least.

What we need to do is verify that $\{x : \phi\}$ exists for each formula ϕ of the language of TST, to ensure that comprehension holds. We do this by induction on the structure of formulas.

$\{x : \neg\phi\}$ is $\{x : \phi\}^c$.

$\{x : \phi \wedge \psi\}$ is $\{x : \phi\} \cap \{x : \psi\}$.

Now we have the much more complex task of analyzing

$$\{t_n : (\forall t_i : \phi(t_1, \ldots, t_n))\}.$$

Choose a type τ' higher than the type τ_i of each t_i. Do this for bound variables as well, and further, in each formula $(\forall t_i : \psi)$ or $(\exists t_i : \psi)$ we require that all occurrences of t_i be free in ψ and the index of t_i be less than the index

of any other variable appearing free in ψ. It should be clear that we can do this without loss of generality.

Where the type of t_i is τ_i, we define x_i as $\iota^{\tau'-\tau_i}(t_i)$: we construct

$$\{t_n : (\forall x_i : \phi(t_1, \ldots, t_n))\}$$

by defining manipulations which allow us to build sets $\{(x_1, \ldots, x_n) : \phi^*(x_1, \ldots, x_n)\}$ in which all the variables are of the same type. We write ϕ^* to suggest that the formula ϕ must be transformed to effect our change of variables: $t_i = t_j$ is equivalent to $x_i = x_j$, and $t_i \in t_j$ is equivalent to $(x_i, x_j) \in [\in]^{\iota^{\tau'-\tau_j}}$ (the reader will see that we use this representation below, though embedded in larger tuples). A quantifier over t_i is equivalent to a quantifier over x_i restricted to $\iota^{\tau'-\tau_i}\text{“}V$.

For any such representation, we have a type signature

$$\iota^{\tau'-\tau_1}\text{“}V \times \ldots \times \iota^{\tau'-\tau_n}\text{“}V.$$

We abbreviate this as τ^*.

The set $\{(x_1, \ldots, x_n) : x_1 = x_n\}$ is obtained as $I^{n-2}_{2,2}(\iota^2\text{“}V) \cap \tau^*$.

The set $\{(x_1, \ldots, x_n) : x_1 = x_i\}$ $(i < n)$ is obtained as $I^{i-2}_{2,3}(I_3(\iota^2\text{“}V)) \cap \tau^*$.

We then can represent the set $\{(x_1, \ldots, x_n) : x_i = x_j\}$ (wlog $i < j$) as $I^{i-1}_{1,n-i+1}(\{(x_1, \ldots, x_{n-i+1}) : x_1 = x_{j-i+1}\}) \cap \tau^*$.

The set $\{(\{x_1\}, x_2) : x_1 \in x_2\}$ has already been defined above as $[\in]$.

The set $\{(x_1, \ldots, x_n) : (\exists uv : x_i = \iota^{k+1}(u) \wedge x_j = \iota^k(v) \wedge u \in v)\}$ is

$$\mathtt{rng}^2(I_3(R^{\iota^{k+2n}}[\in]) \cap \{(x_1, \ldots, x_{n+2}) : x_1 = x_{i+2}\} \cap \{(x_1, \ldots, x_{n+2}) : x_2 = x_{j+2}\}).$$

This handles all representations of membership statements between x_i's in the framework we are using.

The set $\{(x_1, \ldots, x_n) : (\exists x_1 : (x_1, \ldots, x_n) \in R)\}$ is representable as $I_{1,n-1}(\mathtt{rng}(R))$.

This only allows us to quantify over the lowest numbered variable. This is actually sufficient. We have renamed bound variables so that all bound variables with different binders are distinct and in any subformula $(\exists y : \psi)$ y will have lower index than any variable free in ψ. Then the quantified variable might not be x_1, but if it is x_k, we can use the representation $I^k_{1,n-k}(\mathtt{rng}^k(R))$: strip off the first $k-1$ variables, which are dummies already quantified over, quantify over x_k (stripping off one more variable) and put k dummies back, as it were.

This gives sufficient machinery to handle representations of all sentences in the language of TST (or stratified sentences in the language of NF) in the format we are using. However, the final representation of $\{t_n : \phi(t_1, \ldots, t_n)\}$ obtained in this way will be of the form $R = \{(\iota^{\tau'-\tau_1}(t_1), \ldots, \iota^{\tau'-\tau_n}(t_n)) : \phi(t_1 \ldots, t_n)\}$ where some of the t_i's are bound variables (all values of which will appear), and some are parameters. Let P be the set of i such that t_i is a parameter.

The final computation of $\{t_n : \phi(t_1, \ldots, t_n)\}$ will be as

$$\iota^{-(\tau'-\tau_n)}\text{“}\left(\mathtt{rng}^{n-1}\left(R \cap \bigcap_{i \in P}\{(x_1, \ldots, x_n) : x_i = \iota^{\tau'-\tau_i}(t_i)\}\right)\right)$$

where a final bit of computation must be exhibited: $\{(x_1, \ldots, x_n) : x_i = t\}$ is realized as $I_{1,2}^{i-1}(\{\iota^{2(n-i-1)}(t_i)\} \times V) \cap \tau^*$.

We state for the record that we think this is a bad finite axiomatization of NF, or finite template axiomatization of TST: we think the treatment of n-tuples is terribly difficult to work with. But it does work, and the second author chose to verify it, so it merits discussion here. It should be noted that any finite template axiomatization of TST could be used; there is no advantage to this one for the formalization, and there is an oddity, because the one impredicative axiom (the axiom of type lowering) does a lot of extra work, since it is also essential for constructing domains and ranges of functions. In the formalization, the proof of type lowering was in effect divided into the proof of the existence of ι^3"$\mathtt{TL}(x)$, which is predicative, and the proof of the existence of set unions of sets of singletons.

The second author reports that she chose this particular finite axiomatization for verification purposes despite its awkwardness because it uses very few defined concepts, and it is therefore clearer to a reader of the Lean formalization that the proofs indeed say what is required.

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.

3.1 The abstract supertype framework

We first discuss some abstract properties of the types we will construct, and explain the system of "supertypes". 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

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

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].

Proposition 3.1 (the generality of the system of supertypes)**.** Any model of TTT (assuming there are any) is isomorphic to a substructure of a system of supertypes.

Proof. Let M be a model of TTT (more generally, any structure for the language of TTT in which each object outside the base type 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 the problem of constructing a model of TTT is equivalent to the problem of constructing a model of TTT which 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.

Remark 3.2. Notice that if $\alpha >_\tau \beta$ are supertypes , and $x \in \tau^*_\alpha$, $x \cap \tau^*_\beta$ is the extension of x over supertype β. This will be inherited by a scheme of types τ_γ with each $\tau_\gamma \subseteq \tau^*_\gamma$ if an additional condition holds: for $\alpha >_\tau \beta$, we will have for $x \in \tau_\alpha$ that $x \cap \tau^*_\beta$ is the extension of x over supertype β as already noted: for it to be the extension over type β we need the general condition $x \cap \tau^*_\gamma \subseteq \tau_\gamma$ for all $\delta > \gamma$ and $x \in \tau_\delta$.

3.2 Preliminaries of the construction

Now we introduce the notions of our particular construction in this framework.

Definition 3.3 (model parameters)**.** 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 κ be an uncountable regular cardinal (that is, an uncountable regular initial ordinal). Sets of cardinality $< \kappa$ we call *small*. Sets which are not small we may call *large*.

We make no assumption about the relative sizes of κ and λ.

Let μ be a strong limit cardinal with $\kappa < \mu$, $\lambda \leq \mu$ and with the cofinality of μ at least $\mathtt{max}(\kappa, \lambda)$.

Remark 3.4. The minimal model parameters are $\lambda = \omega, \kappa = \omega_1, \mu = \beth_{\omega_1}$.

Definition 3.5 (supertypes)**.** 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 as defined in section 3.2: we now have a definite reference for τ^*_α for $\alpha \in \lambda$. The important part of this definition is that τ^*_{-1} has size μ; the precise form of τ^*_{-1} is chosen to aid with definition 3.12. This type, while important for the model construction, will not be part of our final model of TTT; the types of the final model will be indexed by λ.

Notice that if α, β are type indices, $\alpha \in \beta$ is a convenient short way to say $-1 <_\tau \alpha <_\tau \beta$.

Definition 3.6 (extended type index)**.** A nonempty finite subset of $\lambda \cup \{-1\}$ with minimum element -1 may be termed an *extended type index*. If A is a finite subset of $\lambda \cup \{-1\}$ with at least two elements, A_1 is defined as $A \setminus \{\mathtt{min}(A)\}$, and further A_{n+1} is defined as $(A_n)_1$ where this makes sense: the notation A_2 for $(A_1)_1$ sees use.

Definition 3.7 (atoms, litters and near-litters)**.** We may refer to elements of τ_{-1} 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. This well-ordering is used in only one place in the paper (theorem 4.18), and its use could easily be avoided, but we find its concreteness appealing.

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 that M is *near* N iff $M \sim N$. Note that nearness is an equivalence relation on near-litters.

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

In our model, any near-litter will be the extension of κ-amorphous sets in each type above the base type, and we will use these amorphous sets to limit how much the model knows about the relationship between the different extensions of a model element.

Proposition 3.8. There are exactly μ near-litters.

Proof. A near-litter is determined as the symmetric difference of a litter L (there are μ litters) and a small subset of τ_{-1}. So it is sufficient to show that a set of size μ has no more than μ small subsets. As μ has cofinality at least κ, each small subset of μ is bounded, and so is an element of $\bigcup_{\nu<\mu} \mathcal{P}(\nu)$. But as μ is a strong limit cardinal, each $\mathcal{P}(\nu)$ has size less than μ, so there are only μ bounded subsets of μ; in particular, there are only μ small subsets of μ. □

3.3 Hypotheses for the recursion

The construction of the types τ_α is by a recursion. The initial type τ_{-1} has already been defined. For an ordinal $\alpha < \lambda$ we are assuming that τ_β for $-1 \leq \beta < \alpha$ have already been constructed and satisfy various hypotheses of the recursion. We state hypotheses of the recursion in the following subsections in which the construction of τ_α is described. We list them here as well, but they are likely best understood where they are encountered in the construction. Each of these is enforced for α as well: most of these conditions are explicitly enforced in the course of the construction of τ_α in this section, but that (**I2**) holds for α requires an extensive proof in the next section (theorem 4.33).

This set of inductive hypotheses is not minimal, but the way it is presented has been chosen to match its actual usage in the following sections.

(**I1**) We assume that for $\gamma < \beta < \alpha$, if $x \in \tau_\beta$, $x \cap \tau^*_\gamma \subseteq \tau_\gamma$.

(**I2**) We suppose that each τ_β already constructed is of cardinality μ. Note that we already know that τ_{-1} is of cardinality μ.

(**I3**) We further intimate that for each $x \in \tau_\gamma$, $-1 \leq \gamma < \alpha$, we have defined objects S for which we say that S is a γ-support of x. [This is introduced before the definition of support is actually given in 3.22, but there is no technical obstacle to defining supports earlier if desired].

(**I4**) We define τ^+_β for any type index β as the collection of (x, S) where $x \in \tau_\beta$ and S is a support of x. We assume that we are provided with a well-ordering $\leq^+_\gamma$ of order type μ of τ^+_γ $(-1 \leq \gamma < \alpha)$, by postulating an injection ι^+_γ from τ^+_γ into μ (it does not need to be onto) and defining $x \leq^+_\gamma y$ as $\iota^+_\gamma(x) \leq \iota^+_\gamma(y)$. For any model element x and support S of x, we define $\iota^+_*(x, S)$ as $\iota^+_\gamma(x, S)$, where $x \in \tau_\gamma$. [Hypotheses of the recursion about maps ι^+_γ are stated in (**I11**)].

(**I5**) We provide that for every near litter N and every $\beta < \alpha$, there is a unique element N_β of τ_β such that $N_\beta \cap \tau_{-1} = N$.

(**I6**) We further stipulate that extensionality holds for each $\beta \in \alpha$: that is, for each $\gamma \in \beta$, 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.

(**I7**) We assert that $\iota_*(N) \leq \iota^+_*(N_\delta, S)$ will hold for any near litter N and support S of N_δ, $\delta < \alpha$. This depends on the definition of the position function ι_* which appears below at a convenient point but involves no recursion. (**I7**) appears in (**I11**), but is mentioned earlier here because it is used earlier in the argument.

(**I8**) We presume that all elements of τ_β, $\beta \in \alpha$, are pre-extensional (definition 3.13).

(**I9**) We assume that all elements of τ_β's already constructed are extensional (definition 3.17).

(**I10**) We stipulate that all elements of τ_β ($\beta \in \alpha$) have β-supports. Supports are defined in definition 3.22.

(**I11**) The conditions constraining the choice of functions ι^+_β ($-1 \leq \beta < \alpha$) are

(a) $\iota_*(t) < \iota^+_*(x, S)$ if (t, A) is in the range of S (supports are functions; see definition 3.22) and x is not a typed near-litter.

(b) $\iota_*(N) \leq \iota^+_*(N_\gamma, S)$ for each near-litter N and support S of N_γ (which appeared earlier as (**I7**) for reasons stated there; the notation N_γ is defined in definition 3.9).

We then define $x \leq^+_\beta y$ as $\iota^+_\beta(x) \leq \iota^+_\beta(y)$. We will use the position functions ι^+_* to perform induction along each τ^+_β; these constraints on ι^+_* ensure that objects that a model element depends on (in some suitable sense) are processed before it in the induction.

3.4 Machinery for enforcing extensionality in the model

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 [listed at the end of the previous section]. 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 $x \in \tau_\beta$, $x \cap \tau^*_\gamma \subseteq \tau_\gamma$ (**I1**).

We suppose that each τ_β already constructed ($\beta <_\tau \alpha$) is of cardinality μ. Note that we already know that τ_{-1} is of cardinality μ (**I2**).

We further intimate that for each $x \in \tau_\gamma$, $-1 \leq \gamma < \alpha$, we have defined objects S for which we say that S is a support of x (**I3**). The definition of supports will be given in definition 3.22. 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, from which it follows that τ_γ^+ is of cardinality μ, since there are μ supports (as we will see when supports are defined).

We also provide a well-ordering $\leq_\gamma^+$ of order type μ of τ_γ^+ $(-1 \leq \gamma < \alpha)$, by postulating an injection ι_γ^+ from τ_γ^+ into μ (it does not need to be onto) and defining $x \leq_\gamma^+ y$ as $\iota_\gamma^+(x) \leq \iota_\gamma^+(y)$ (**I4**). There are some hypotheses of the recursion about maps ι_γ^+ which are stated below. For any model element x and support S of x, we define $\iota_*^+(x, S)$ as $\iota_\gamma^+(x, S)$, where $x \in \tau_\gamma$.

We provide that for every near litter N and every $\beta < \alpha$, there is a unique element N_β of τ_β such that $N_\beta \cap \tau_{-1} = N$ ((**I5**): we will quite shortly give a precise description of all extensions of this object).

Definition 3.9 (typed objects)**.** If X is a subset of τ_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 N is a near-litter we refer to N_β as a *typed near-litter* (of type β).

We further stipulate (**I6**) that extensionality holds for each $\beta \in \alpha$: that is, for each $\gamma \in \beta$, 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.

*It is interesting to note that the Lean formalisation does not use (**I6**) until the counting argument; in particular, it is not needed to construct τ_α at stage α.*

Definition 3.10 (position function for near-litters and singletons of atoms)**.** We posit a bijection ι_* from the set of all near-litters and singletons of atoms to μ with the following properties:

1. $\iota_*(L) < \iota_*(\{a\})$ if $a \in L$ and L is a litter.
2. $\iota_*(N^\circ) < \iota_*(N)$ if N is a near-litter which is not a litter
3. $\iota_*(\{a\}) < \iota_*(N)$ if $a \in N \Delta N^\circ$

The function may be constructed directly: first choose an ordering of type μ on litters, then put singletons of atoms directly after the litter they are inside (and before all later litters), then put near-litters after the corresponding litter and all singletons of atoms in the symmetric difference, then map each near-litter or singleton of an atom to its position in the resulting order. We don't run out of room before finishing the construction because μ has cofinality at least κ. We prove more formally that there are such functions just below.

Lemma 3.11 (position function generator)**.** If X is a set of size at most μ and $f : X \to \mathcal{P}(\mu)$ is such that each $f(x)$ is bounded, then there is an injection $\iota : X \to \mu$ such that $\iota(x) \notin f(x)$.

Proof. Choose a well-ordering $<_X$ of X of type at most μ.

At each stage x, assume we have already constructed $\iota(y)$ for $y <_X x$.

We then define

$$\iota(x) = \min\left(\mu \setminus f(x) \setminus \{\iota(y) : y < x\}\right).$$

Note that as $f(x)$ is bounded in μ, and μ has initial order type, the set $\mu \setminus f(x)$ is of cardinality μ.

Since $\{\iota(y) : y < x\}$ has cardinality less than μ, the difference must be nonempty, so the definition given always succeeds.

□

We now show how to construct a position function for near-litters and singletons of atoms satisfying the given constraints.

Proof. First, let ι_0 be any injection from the set of litters into μ.

Next, use lemma 3.11, with X taken to be the collection of all singletons of atoms, setting $f(\{a\}) = \{\nu : \nu \leq \iota_0(L)\}$ where L is the unique litter such that $a \in L$, to create ι_1.

Finally, use lemma 3.11 again, with X taken to the collection of all near-litters, setting

$$f(N) = \{\nu : \nu \leq \iota_0(N^\circ)\} \cup \bigcup_{a \in N \Delta N^\circ} \{\nu : \nu \leq \iota_1(a)\},$$

to obtain ι_2.

Now, we define a map ι from near-litters and singletons of atoms to $3 \times \mu$ by

$$g(x) = \begin{cases} (0, \iota_0(x)) & \text{if } x \text{ is a litter} \\ (1, \iota_1(x)) & \text{if } x \text{ is a singleton} \\ (2, \iota_2(x)) & \text{otherwise} \end{cases}$$

We can then define $\iota_*(x)$ to be the position of $g(x)$ in the restriction of the lexicographic order on $3 \times \mu$ to the image of g (which is of order type μ).

The resulting function ι_* is then clearly a bijection, and has the required properties by construction, noting that near-litters that happen to be litters will automatically satisfy the third condition.

□

An additional property involving ι_* which is enforced by inductive hypotheses explained later about the maps ι^+_β is that $\iota_*(N) \leq \iota^+_*(N_\delta, S)$ will hold for any near litter N and support S of N_δ (**I7**). It must be noted here because it is shortly used.

Definition 3.12 (f maps, crucially important)**.** [5] We define for each type index β less than α and each ordinal γ distinct from β 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 the set of litters, with range included in $\{L_{\nu,\beta,\gamma} : \nu < \mu\}$ to ensure that distinct f maps have disjoint ranges.

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, where $\iota_*(L)$ is minimal such that

1. $L \in \{L_{\nu,\beta,\gamma} : \nu < \mu\}$,
2. $\iota_*^+(x) < \iota_*(L)$ [and so for any $N \sim L$, $\iota_*^+(x) < \iota_*(N)$, and for any $z \in L$, $\iota_*^+(x) < \iota_*(\{z\})$],
3. and for any $y <_\beta^+ x$, $f_{\beta,\gamma}(y) \neq L$.

Note that the ranges of distinct f maps are disjoint sets.

Definition 3.13 (pre-extensional)**.** We define the notion of *pre-extensional* element of τ_β^* ($-1 < \beta \leq \alpha$). 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 $x \cap \tau_\delta$ ($\delta \in \beta$) 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, $\beta \in \alpha$, are pre-extensional (**I8**).

*This hypothesis (**I8**) is not used in Lean. The objects we construct at stage α satisfy this property, but we don't use this outside of (Lean's equivalent of) section 3, and we never need to know that lower-type objects satisfy it. I believe that this hypothesis is only tacitly used in the form that allowable permutations preserve (pre-)extensionality.*

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 now verify that the order conditions in the definition of $f_{\beta,\gamma}$ ensure that distinguished extensions are unique.

Proposition 3.14. For any $x \in \tau_\beta$ there is only one set $x \cap \tau_\gamma$ which is a distinguished extension of x.

[5] The use of model elements with support rather than simply model elements as domain elements of the f maps is a substantial contribution of the second author to the mathematics of the paper, above simply verifying the work of the first author. The proof could be carried out without this, but it is much easier to present with this refinement. There are other ways in which the second author has contributed to the mathematics, but this one is especially worthy of note.

Proof. If any $x \cap \tau_\gamma$ ($\gamma \in \beta$) is empty or if $x \cap \tau_{-1}$ is not empty, $x \cap \tau_{-1}$ is the unique distinguished extension (if it is empty of course it coincides with all the other extensions).

So, what remains is the case of x with $x \cap \tau_{-1}$ empty and each $x \cap \tau_\gamma$ nonempty for $\gamma < \beta$.

If $x \cap \tau_\gamma$ is an extension of x and $a \in x \cap \tau_\gamma$ is not of the form N_γ for a near-litter N, then $x \cap \tau_\gamma$ must be the unique distinguished extension: the reason for this is that for any distinguished extension, all elements $b \in x \cap \tau_\delta$ of other extensions must be N_δ's.

So we are down to the case where $x \cap \tau_\delta$ is nonempty for $\delta \neq -1$, $x \cap \tau_{-1}$ is empty, and each element of any $x \cap \tau_\delta$ is of the form N_δ where N is a near-litter. Let $x \cap \tau_\gamma$ be a distinguished extension. For any $M_\delta \in x$ with $\delta \neq \gamma$, we have $M \sim f_{\gamma,\delta}(P_\gamma, Q)$ for some near-litter P and support Q of P_γ. Then we have $\iota_*(M) > \iota_*^+(P_\gamma, Q) \geq \iota_*(P)$ by (**I7**). Let $N_\gamma \in x$ be chosen so that $\iota_*(N)$ is minimal. It follows that for any $\delta \neq \gamma$ and $M_\delta \in x$ we have $\iota_*(M) > \iota_*(N)$, from which it is evident that we cannot have two distinct distinguished extensions: if δ were the index of another distinguished extension, and $\iota_*(M)$ were chosen minimal so that $M_\delta \in x$, it would follow that $\iota_*(M) < \iota_*(N)$ as above but also that $\iota_*(N) < \iota_*(M)$, which is absurd. □

Definition 3.15 (A map)**.** 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 x 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 $A^{-1}(x)$ for this set if it exists, defining a very partial function A^{-1} on nonempty subsets of types.

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

Proposition 3.16. No subset of a type has infinitely many iterated images under A^{-1}.

Proof. Let a be a subset of type γ for which $A^{-1}(a)$ exists.

Since $A^{-1}(a)$ exists, every element of a is of the form N_γ where N is a near-litter. Choose $N_\gamma \in A^{-1}(a)$ such that $\iota_*(N)$ is minimal. Note that N is in fact a litter by the constraints in definition 3.10.

Let $b = A^{-1}(a)$: we have $a = A_\gamma(b)$, where b is a subset of some τ_δ with $\delta \neq \gamma$. In particular, $N = f_{\delta,\gamma}(u, U)$ for some $u \in \tau_\delta$ and U a support of u. If $u = M_\delta$ we can further state that $\iota_*(N) > \iota_*^+(M_\delta, U) \geq \iota_*(M)$: if b itself has an image under A^{-1}, the minimum value of ordinals $\iota_*(M)$ for $M_\delta \in b$ will be less than the minimum value of ordinals $\iota_*(N)$ for $N_\gamma \in a$, which establishes that there is an ordinal parameter determined by a nonempty subset of a type which

decreases strictly when A^{-1} is applied (if it is applicable), and so no nonempty subset of a type can have infinitely many iterated preimages under A^{-1}. This is a rephrasing of an argument which occurred above in the discussion of the uniqueness of distinguished extensions. □

Definition 3.17 (extensional)**.** 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}.[6]

Proposition 3.18 (extensionality)**.** Two extensional model elements with any common extension (over a type other than τ_{-1}) are equal.

Proof. If two extensional model elements have an empty extension (over a type other than τ_{-1}) 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 the two elements 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 assume that all elements of τ_β's already constructed are extensional (**I9**). This completes the mechanism for enforcement of extensionality in the structure we are defining.

*Again, the Lean formalisation only uses this for the construction of τ_α, and nowhere else. Of course, we need to remember the conclusion of proposition 3.18 (namely, (**I6**)), but that is all that is needed. In my opinion, the better way to phrase this part is that extensional elements at type α are candidates for inclusion in τ_α since they satisfy (**I6**), and remove any mention of (pre-)extensional element from the inductive hypotheses; this also means we can specialise the definitions of (pre-)extensional elements to τ_α.*

3.5 Allowable permutations and supports

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 (**I2**). It is certainly not a sufficient restriction to require elements of τ_α to be extensional: we will require a further symmetry condition.

We define classes of permutations of our structures.

[6] We do know that we are carefully, explicitly, spelling out a construction which looks very much like the construction of the bijection in the Cantor-Schröder-Bernstein theorem. But the details of the maps involved are used, so everything must be spelled out.

Definition 3.19 (structural permutation)**.** A -1-*structural permutation* is a permutation of $\tau^*_{-1} = \tau_{-1}$.

A β-*structural permutation* ($-1 < \beta \leq \alpha$) is a permutation π of τ^*_β such that for each type $\gamma < \beta$ there is a γ-structural permutation π_γ such that $\pi(x) \cap \tau^*_\gamma = \pi_\gamma\text{“}(x \cap \tau^*_\gamma)$ for any $x \in \tau^*_\beta$.

The maps π_γ are referred to as *derivatives* of π. If π is a -1-structural permutation, π_{-1} may be taken to denote π itself.

More generally, for any finite subset A of $\lambda \cup \{-1\}$ with maximum α, define π_A as $(\pi_{A\setminus\{\mathtt{min}(A)\}})_{\mathtt{min}(A)}$ and $\pi_{\{\alpha\}} = \pi$. The maps π_A may be referred to as iterated derivatives of π.[7] It should be clear that a structural permutation is exactly determined by its iterated derivatives which are -1-structural.

Where π is a β-structural permutation with $\beta > -1$, we define π^+ as π_{-1}. This is occasionally useful to reduce notational clutter.

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

Definition 3.20 (allowable permutation)**.** A -1-*allowable permutation* is a permutation π of τ_{-1} such that for any near-litter N, $\pi\text{“}N$ is a near-litter.

A β-*allowable permutation* ($\beta \leq \alpha$) is a β-structural permutation, each of whose derivatives π_γ is a γ-allowable permutation (and satisfies the condition that $\pi_\gamma\text{“}\tau_\gamma = \tau_\gamma$) and which satisfies a coherence condition relating the f maps and derivatives of the permutation: for suitable $\gamma, \delta < \beta$,

$$f_{\gamma,\delta}(\pi_\gamma(x), \pi_\gamma[S]) \sim \pi^+_\delta\text{“}f_{\gamma,\delta}(x, S).$$

(where the action of allowable permutations on supports will be defined shortly). This coherence condition is motivated in remark 3.26.

Note that a β-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.

Definition 3.21 (support condition)**.** A β-*support condition* ($-1 \leq \beta \leq \alpha$) is defined as a pair (x, A), where

1. A is an extended type index with maximum β. (Recall that an extended type index has minimum -1).

2. and $x \subseteq \tau_{-1}$ is either a singleton or a near-litter, and must be a singleton in the case $\beta = -1$.

[7] There is a silly notational point here: we might want to suppose π_A and π_α to be essentially distinguished in some way we do not actually implement (for example by type face) in order to prevent confusion of $\pi_{\{0\}}$ with π_1. A similar problem exists due to the identification of finite ordinals with finite sets of ordinals. However, it can also be noted that $\pi_{\{0,\ldots,n\}}$ and π_{n+1} cannot make sense for the same allowable permutation π, so we think this is harmless.

Definition 3.22 (support)**.** Where $-1 \leq \beta \leq \alpha$, a *$\beta$-support* is defined as a function S from a small ordinal to β-support conditions.

We may write S_δ intead of $S(\delta)$.

Remark 3.23. -1-supports behave somewhat differently from β-supports for $-1 < \beta$, and we consider the notion of -1-support to be merely a technical convenience, often making the statements of definitions and lemmas more uniform. The condition that -1-support conditions cannot contain near-litters is used exactly once, in the proof of proposition 4.35.

Definition 3.24 (operations on supports)**.** We define various operations to manipulate supports.

1. For any support condition (x, B) we define $(x, B)^{\uparrow A}$ as $(x, B \cup A)$ if all elements of the set A dominate all elements of the set B. Further, if S is a support, we define $S^{\uparrow A}$ so that $(S^{\uparrow A})_\epsilon = (S_\epsilon)^{\uparrow A}$. By an abuse of notation we may write $(x, B)^{\uparrow \beta}$ or $S^{\uparrow \beta}$ where β is an ordinal for $(x, B)^{\uparrow \{\beta\}}$ or $S^{\uparrow \{\beta\}}$.

2. For any supports S and T we denote by $S+T$ a support which consists of S, followed by T: what this means is that $(S+T)_\epsilon = S(\epsilon)$ [which we write S_ϵ] for ϵ in the domain of S, $(S+T)_{\mathtt{dom}(S)+\epsilon} = T_\epsilon$ for ϵ in the domain of T.

3. We define the action of a β-allowable permutation π on a β-support S: if $S(\delta) = (x, A)$, $\pi[S](\delta) = (\pi_A\text{“}x, A)$.

4. An element x of τ^*_β *has β-support* S iff for every β-allowable permutation π, if $\pi[S] = S$ then $\pi(x) = x$. We say that an element x of τ^*_β which has a β-support is *β-symmetric.*

Remark 3.25 (counting supports). It is straightforward to observe that there are μ β-supports for $\beta \leq \alpha$: there are μ atoms, μ near-litters, and $< \mu$ finite subsets of $\beta + 1 < \lambda \leq \mu$ (all type indices involved in a β-support are $\leq \beta$); thus the set of β-support conditions (which we will call $\mathtt{SC}$ for the moment) is of size μ; note that the set $\kappa \times \mathtt{SC}$ is of cardinality μ and each β-support is a small subset of $\kappa \times \mathtt{SC}$, and so, as we have already seen than sets of size μ have μ small subsets, it follows that there are no more than μ β-supports. Thus τ^+_β is already known to be of size μ for $\beta < \alpha$.

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

Remark 3.26 (motivation of the coherence condition in definition 3.20). The motivation for this is that we need β-allowable permutations ($\beta \leq \alpha$) to send extensional elements of supertypes to extensional elements. Suppose $x \in \tau_\beta$ and $x \cap \tau_\gamma = \{b\}$. If x is extensional, this has to be the distinguished extension of x. For any $\delta \in \beta \setminus \{\gamma\}$, it follows that $x \cap \tau_\delta$ is the set of all N_δ such that $N \sim f_{\gamma,\delta}(b, S)$ for some support S of b. This tells us that a β-allowable permutation π, such that $\pi(x)$ has γ-extension $\{\pi_\gamma(b)\}$, must have the δ-extension of $\pi(x)$ equal to

$$\pi_\delta\text{“}\{N_\delta : \exists S : N \sim f_{\gamma,\delta}(b, S)\}$$

but must also have its δ-extension equal to

$$\{N_\delta : \exists S : N \sim f_{\gamma,\delta}(\pi_\gamma(b), S)\}.$$

This tells us that

$$\pi_\delta(f_{\gamma,\delta}(b,S)_\delta) \in \{N_\delta : (\exists T : N \sim f_{\gamma,\delta}(\pi_\gamma(b), T))\}$$

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 $\pi_\beta(b)$.

Proposition 3.27 (allowable permutations preserve extensionality)**.** Allowable permutations map extensional elements of supertypes to extensional elements.

Proof. Recall that we defined $A_\delta(a)$ in definition 3.15 as

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

If π is allowable of suitable index, $\pi_\delta\text{“}A_\delta(a) = A_\delta(\pi_\gamma\text{“}a)$ 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

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

So any element of $\pi_\delta\text{“}A_\delta(a)$ is in $A_\delta(\pi_\gamma\text{“}a)$.

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

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

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 $A^{-1}(\pi_\gamma\text{“}a)$ exists and is in τ_δ, and moreover $A^{-1}(\pi_\gamma\text{“}a)$ is equal to $\pi_\delta\text{“}A^{-1}(a)$ 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. □

3.6 Model elements defined

Definition 3.28 (model elements)**.** We stipulate that all elements of τ_β ($\beta \in \alpha$) have β-supports [enforcing (**I10**)], 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.

Note that an image of an element of τ_β ($\beta \leq \alpha$) under a β-allowable permutation will belong to τ_β, because supportedness and extensionality are preserved by allowable permutations.

The definition explicitly enforces (**I1**), (**I3**), (**I5**) (a near-litter obviously has a support), (**I6**), (**I8**), (**I9**), (**I10**) for subsequent stages of the construction.

We still have to prove that the cardinality of τ_α, and so of τ_α^+, is μ, to show that the construction works (verification of (**I2**) for subsequent stages is in the next section). However, we can show now that given (**I2**), we can satisfy the remaining hypotheses (**I4**), (**I7**), (**I11**).

Remark 3.29 (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, whose range is obtained from the union of the ranges of the supports of the elements of X by replacing each element (u, B) of the union of the ranges 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}).

Definition 3.30 (position functions)**.** The conditions constraining the choice of functions ι_β^+ (enforcing (**I4**), (**I7**), (**I11**)) are

1. $\iota_*(t) < \iota_\beta^+(x, S)$ if (t, A) is in the range of S and x is not a typed near-litter (**I11**).
2. $\iota_*(N) \leq \iota_\beta^+(N_\beta, S)$ for any near-litter N and support S of N (**I7**).

We then define $x \leq_\beta^+ y$ as $\iota_\beta^+(x) \leq \iota_\beta^+(y)$.

There is no difficulty in satisfying these constraints given that τ_β^+ has size μ (**I2**) as there are only a small set of constraints on any particular value of ι_*^+ and μ is of cofinality at least κ.

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 4.1 (-1-approximation)**.** A -1-*approximation* is a function ψ such that:

1. The domain and image of ψ are the same and ψ is injective.
2. Each domain element of ψ is either an atom or a litter, and moreover, ψ maps atoms to atoms and litters to litters.
3. For each litter L, ψ and ψ^{-1} are defined on only a small collection of atoms $a \in L$.

We will associate a partial function ψ^* on atoms to each -1-approximation ψ. The action of ψ on atoms will agree with the action of ψ^*, and the action of ψ on litters will agree with the pointwise action of ψ^* only up to nearness.

Definition 4.2. If ψ is a -1-approximation, we define the partial function ψ^* by:

1. If a is an atom and $a \in \mathtt{dom}(\psi)$, then $\psi^*(a) = \psi(a)$.
2. If a is an atom with $a \notin \mathtt{dom}(\psi)$ but $a \in L$ and $L \in \mathtt{dom}(\psi)$, then

$$\psi^*(a) = \pi_{M,N}(a), \text{ where } M = L \setminus \mathtt{dom}(\psi) \text{ and } N = \psi(L) \setminus \mathtt{dom}(\psi)$$

where for any co-small subsets of litters M, N, the map $\pi_{M,N}$ is the unique map ρ from M to N such that for any $x, y \in M$,

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

$\pi_{M,N}$ is the unique map from M onto N which is strictly increasing in the order determined by fourth projections of atoms. Notice that $\pi_{M,N} \circ \pi_{L,M} = \pi_{L,N}$ will hold if L, M, N are co-small subsets of litters, and $\pi_{L,M} \circ \pi_{M,L} = \pi_{L,L}$ which is the identity map on L, under the same conditions.[8]

Remark 4.3. If N is a near-litter and $N \subseteq \mathtt{dom}(\psi^*)$, then $N^\circ \in \mathtt{dom}(\psi)$, and additionally $\psi^* \text{“} N \sim \psi(N^\circ)$. The converse is not true: $N^\circ \in \mathtt{dom}(\psi)$ does not imply that $N \subseteq \mathtt{dom}(\psi^*)$, but it does imply that $N^\circ \subseteq \mathtt{dom}(\psi^*)$.

Note that if n is any integer, ψ^n is also a -1-approximation with the same domain, where we take the convention that ψ^0 is the identity map on $\mathtt{dom}(\psi)$. Using the condition $\pi_{M,N} \circ \pi_{L,M} = \pi_{L,N}$, we obtain the equation $(\psi^n)^* = (\psi^*)^n$.

[8] 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 like the concreteness of this approach.

From now on, we avoid using the action of ψ directly where possible, and instead use ψ^*.

Remark 4.4. ψ^* is a permutation of atoms. If it is defined on all of τ_{-1}, it is a -1-allowable permutation.

Definition 4.5 (extension)**.** We define a partial order on -1-approximations by setting $\psi \preceq_{-1} \chi$ if $\psi \subseteq \chi$ and $\mathtt{dom}(\psi) \cap \tau_{-1} = \mathtt{dom}(\chi) \cap \tau_{-1}$. That is, χ may define images for more litters than ψ, but may not define images for any new atoms. If $\psi \preceq_{-1} \chi$, we may call χ an *extension* of ψ. Note that if $\psi \preceq_{-1} \chi$, then $\psi^* \subseteq \chi^*$.

Lemma 4.6 (adding orbits)**.** Let ψ be a -1-approximation, and let $(L_n)_{n\in\mathbb{Z}}$ be litters such that

$$\{\langle L_n, L_{n+1}\rangle : n \in \mathbb{Z}\}$$

is a bijection, and ψ^* is not defined at any of the L_n. Then ψ has an extension χ where $\chi^*\text{“}L_n \sim L_{n+1}$ for each n, and all near-litters contained in the domain of χ^* but not contained in the domain of ψ^* are near some L_n.

Proof. Define

$$\chi = \psi \cup \{\langle L_n, L_{n+1}\rangle : n \in \mathbb{Z}\};$$

it is easy to verify all of the required conditions. □

It is not necessary that all of the L_n be distinct. This lemma therefore allows us to add single orbits of litters of any order to -1-approximations.

Definition 4.7 (approximation, in general)**.** If $-1 < \beta$, a β-approximation is defined as a function with $\{-1\} \cup \beta$ as domain such that $\psi(\gamma)$ is a γ-approximation for each γ in the domain. We write ψ_γ instead of $\psi(\gamma)$.

We define some operations on approximations.

1. If A is a finite subset of $\lambda \cup \{-1\}$ with maximum element β, we define the derivative of ψ along A in the following way: $\psi_A = (\psi_{A\setminus\{\mathtt{min}(A)\}})_{\mathtt{min}(A)}$ and $\psi_{\{\beta\}} = \psi$.

   If $\beta = -1$, construe ψ_{-1} as ψ.

2. A β-approximation ψ acts on a support S by

   $$(\psi^*[S])_\delta = (\psi^*_A\text{“}x, A) \text{ where } S_\delta = (x, A)$$

   whenever this is defined for each $\delta \in \mathtt{dom}(S)$.

3. We define the partial order $\preceq_\beta$ on β-approximations by defining $\psi \preceq_\beta \chi$ whenever $\psi_\gamma \preceq_\gamma \chi_\gamma$ for all $\gamma < \beta$, and define an *extension* of a β-approximation ψ as an approximation $\chi \succeq_\beta \psi$.

Definition 4.8 (flexibility)**.** A near-litter N is *A-flexible*, where A is an extended type index, if $|A| \leq 2$ or N° is not in the range of any $f_{\gamma,\mathtt{min}(A_1)}$ for $-1 \leq \gamma < \mathtt{min}(A_2)$.

Definition 4.9 (coherent)**.** Let $-1 \leq \beta \leq \alpha$. We say that a β-approximation ψ is *coherent* (c.f. the coherence condition on allowable permutations from definition 3.20) if:

1. If L is A-flexible for some A, then $\psi_A^* \text{“} L$ is also A-flexible.
2. If A is an extended type index with maximum β and $\mathtt{min}(A_1) = \gamma < \beta$, and $\delta < \mathtt{min}(A_2)$ and $(x, S) \in \tau_\delta^+$ are such that $f_{\delta,\gamma}(x, S) \subseteq \mathtt{dom}(\psi_A^*)$, then there is some δ-allowable permutation π such that

$$(\psi_{A_2})_\delta^*[S] = \pi[S]$$

and additionally that

$$\psi_A^* \text{“} f_{\delta,\gamma}(x, S) \sim f_{\delta,\gamma}(\pi(x), \pi[S])$$

(and hence all δ-allowable permutations π satisfying the given hypotheses also satisfy the stated coherence condition).

Remark 4.10. If ψ is coherent, then ψ^n is coherent for any integer n, and ψ_γ is coherent for any $\gamma < \beta$. Every -1-approximation is coherent.

Definition 4.11 (approximate)**.** A -1-approximation ψ is said to *approximate* a -1-allowable permutation π if $\psi^* \subseteq \pi$. If $-1 < \beta$, a β-approximation ψ is said to *approximate* a β-allowable permutation π if for each $\gamma < \beta$, ψ_γ approximates π_γ.

Remark 4.12. If $\psi \preceq_\beta \chi$ and χ approximates π, then ψ approximates π.

Definition 4.13 (freedom of action)**.** We say that *freedom of action* holds at some type index β if every coherent β-approximation ψ approximates some β-allowable permutation π.

Remark 4.14. If ψ is a coherent approximation such that ψ_A is defined on all litters (or equivalently, ψ_A^* is defined on all atoms) for all A containing -1, then it is easy to see that ψ approximates a unique allowable permutation π, and π is given by $\pi_A = \psi_A^*$. So to prove that freedom of action holds at level β, it suffices by remark 4.12 to show that every coherent β-approximation has a coherent extension χ such that χ_A is defined on all litters for all A containing -1.

Lemma 4.15. Let ψ be a coherent β-approximation. Then ψ admits a coherent extension χ such that for each extended type index A with maximum β, χ_A is defined on all A-flexible litters.

Proof. The construction

$$\chi_A = \psi_A \cup \{\langle L, L\rangle : L \notin \mathtt{dom}(\psi_A^*), L \text{ is } A\text{-flexible}\}$$

suffices. □

Remark 4.16. This lemma shows in particular that freedom of action holds at type -1. Indeed, suppose that ψ is a (coherent) -1-approximation and χ is an extension as above. By remark 4.4, χ^* is a -1-allowable permutation as all near-litters are $\{-1\}$-flexible, and χ clearly approximates it. But $\psi \preceq_{-1} \chi$, so ψ also approximates χ^* (remark 4.12).

Lemma 4.17. Let ψ be a coherent β-approximation. Let A be an extended type index with maximum β, and let $\gamma = \mathtt{min}(A_1)$ and $\delta < \mathtt{min}(A_2)$. Let $(x, S) \in \tau_\delta^+$ be such that $(\psi_{A_2})^*_\delta[S]$ is defined.

Then if freedom of action holds at type level δ, there is a coherent extension $\chi \succeq_\beta \psi$ such that $\chi^*_A \text{“} f_{\delta,\gamma}(x, S)$ is defined.

Proof. The δ-approximation $(\psi_{A_2})_\delta$ is coherent (remark 4.10), so it approximates some δ-allowable permutation π by freedom of action. In particular, $((\psi^n_{A_2})_\delta)^*[S] = \pi^n[S]$ for every integer n. We intend to add the orbit

$$f_{\delta,\gamma}(\pi^n(x), \pi^n[S]) \mapsto f_{\delta,\gamma}(\pi^{n+1}(x), \pi^{n+1}[S])$$

to ψ^*_A, at least up to nearness.

Suppose that there is some n such that ψ^*_A is defined on $f_{\delta,\gamma}(\pi^n(x), \pi^n[S])$. We have $((\psi^n_{A_2})_\delta)^*[S] = \pi^n[S]$, so $((\psi^{-n}_{A_2})_\delta)^*[\pi^n[S]] = S$. Therefore, as ψ^{-n} is coherent (also by remark 4.10), we obtain

$$(\psi^{-n})^*_A \text{“} f_{\delta,\gamma}(\pi^n(x), \pi^n[S]) \sim f_{\delta,\gamma}(x, S).$$

Therefore,

$$f_{\delta,\gamma}(x, S) \subseteq \mathtt{dom}(\psi^*_A).$$

So we do not need to extend ψ; we are already done.

Otherwise, we can use lemma 4.6 to extend ψ to an approximation χ in which

$$\chi^*_A \text{“} f_{\delta,\gamma}(\pi^n(x), \pi^n[S]) \sim f_{\delta,\gamma}(\pi^{n+1}(x), \pi^{n+1}[S])$$

for each integer n. It is easy to check that this χ is coherent. □

Theorem 4.18 (Freedom of Action)**.** Freedom of action holds at all type indices $\beta \leq \alpha$.[9]

Proof. By induction, we may assume freedom of action holds at all levels $\delta < \beta$. Let ψ be a coherent β-approximation, and use Zorn's lemma to extend ψ to a maximal coherent extension χ; this step uses the fact that coherence is preserved under suprema of chains of approximations.

Suppose that χ^*_A is not defined on all litters for some A. Let L be the litter with minimal position $\iota_*(L)$ such that there is an extended type index A with maximum element β such that $L \nsubseteq \mathtt{dom}(\psi^*_A)$.

[9]It should be noted that the second author has entirely reorganized and rewritten this proof, and indeed the entire section on Freedom of Action. The underlying mathematics is the same but the presentation is much cleaner.

Suppose that L is A-flexible. By lemma 4.15, χ admits a coherent extension φ such that $L \subseteq \mathtt{dom}(\varphi^*_A)$. This contradicts maximality of χ.

Now suppose that L is not A-flexible. Writing γ for the minimum element of A, there are $\delta < \mathtt{min}(A_1)$ and $(x, S) \in \tau^*_\delta$ such that $L = f_{\delta,\gamma}(x, S)$.

We claim that $(\psi_{A_2})^*_\delta[S]$ is defined; this will then give a contradiction by lemma 4.17. To show this, we must prove that if $(t, B) \in \mathtt{rng}(S)$, then $\psi^*_A\text{“}t$ is defined. By definition 3.12, we have $\iota^+_*(x, S) < \iota_*(L)$, and by (**I11**), $\iota_*(t) \leq \iota^+_*(x, S)$. If t is a near litter or singleton of an atom and $t \subset M$ for a litter M, then by definition 3.10, we obtain $\iota_*(M) < \iota_*(t)$, and therefore that $\psi^*_B\text{“}t$ is defined by minimality of the position of L. Alternatively, if t is a near-litter, then if M is any litter such that $M \cap t \neq \emptyset$, definition 3.10 again implies that $\iota_*(M) < \iota_*(t)$, so $\psi^*_B\text{“}M$ is defined. Combining all such litters, we conclude that $\psi^*_B\text{“}t$ is defined.

Therefore χ^*_A must be defined on all litters for all A. By remark 4.14, this concludes the proof: ψ approximates the allowable permutation π given by $\pi^*_A = \chi^*_A$. □

Remark 4.19. The use of Zorn's lemma in the previous proof is merely a technical convenience. Its use can be excised by instead computing the value of χ^*_A at each atom directly, under the inductive hypothesis that its value at each atom at an earlier position was already computed for all A.

In fact, all of the proofs of this subsection can be phrased in such a way that the axiom of choice is not invoked. In particular, the coherent extension defined in 4.17 is uniquely determined: all choices of π yield the same extension χ. This means that every coherent approximation defined on all flexible litters has a unique coherent extension defined on all litters (but we will not use this fact).

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 (verification of (**I2**) for subsequent stages of the construction is thus completed).

Definition 4.20 (interference)**.** Let $x, y \subseteq \tau_{-1}$. Their *interference* is defined to be the union of the small elements of $\{x \Delta y, x \cap y\}$, which is a small subset of τ_{-1}.

Thus the interference between two near-litters M and N is either $M \Delta N$ or $M \cap N$, whichever is small.

Definition 4.21 (strong support)**.** A β-support S is called *strong* if it satisfies the additional properties that

1. if (x, A) and (y, A) are in the range of S, then $(\{z\}, A) \in S$ for all atoms z in the interference of x and y,
2. and for each ϵ in the domain of S, if $S_\epsilon = (x, A)$, and $x^\circ = f_{\gamma,\delta}(y, T)$, then the range of $T^{\uparrow A_2}$ is a subset of the range of $S\lceil\epsilon$: supports appearing in inverse images under f of litters which are near the first projections of an element of the support have a type-raised copy (mod reindexing) appearing in the support before that item.

Remark 4.22. It should be evident that if π is a β-allowable permutation and S is a strong β-support, $\pi[S]$ is also a strong β-support.

Remark 4.23. *TODO: Re-read this, maybe convert to a proposition? Note that in Lean we only need that S has range a subset of the range of a strong support. (This comment is left visible deliberately to encourage the revision suggested).* Each support S is the terminal segment of a strong support. Such a strong support can be constructed by prefixing to S, for each $T^{\uparrow A_2}$ such that for some ϵ in the domain of S, $S_\epsilon = (x, A)$, and $x^\circ = f_{\gamma,\delta}(y, T)$, a strong support with $T^{\uparrow A_2}$ as a terminal segment, which will be obtainable as $U^{\uparrow A_2}$ where U is a strong support with T as a terminal segment, which exists by the inductive hypothesis that this is true for supports with lower type index than S, followed by all atomic items which must be added to satisfy the first condition in the definition of strong support. In fact, we can define a canonical strong support of which S is a terminal segment by stipulating that the $U^{\uparrow A_2}$'s are added (each one in turn between the preceding ones and S) in the order in which the correlated S_ϵ's appear in S and that U is the canonical downward extension of T in each case, and that the atomic items are added in the order of their images under ι^+_*, with some fixed well-ordering of extended type indices used to resolve order of items with the same value under ι^+_*.

Definition 4.24 (coding function)**.** 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 4.25 (designated support)**.** For each ordinal γ, and for each orbit in τ_γ under allowable permutations, choose x in the orbit (the designated element of the orbit), choose a strong support S of x,and for every other y in the orbit, choose a designated allowable permutation π_y such that $\pi_y(x) = y$, and define the *designated support* of y to be $\pi_y[S]$.

Definition 4.26 (specification)**.** A *η-specification* S^* of a η-support S is a function with the same domain as S. We use the notation S^*_ϵ for $S^*(\epsilon)$.

1. If S_ϵ is $(\{x\}, A)$, where $\beta = \mathtt{min}(A_1)$, then S^*_ϵ is $(0, \beta, \Sigma, A)$ where Σ is the set of all δ such that $A = \pi_2(S_\delta)$ and $x \in \pi_1(S_\delta)$ (this captures identical atoms and near litters containing the given atom)

2. If S_ϵ is (N, A) and N is a near-litter, where $\beta = \mathtt{min}(A_1)$, and either $|A| \leq 2$ or N° is not in the range of any $f_{\gamma,\beta}$ for $\gamma < \mathtt{min}(A_2)$ (that is, N is A-flexible), then S^*_ϵ is $(1, \beta, \Sigma, A)$, where Σ is the set of all δ such that $\pi_2(S_\delta) = A$ and $\pi_1(S_\delta) \sim \pi_1(S_\epsilon)$.

3. If S_ϵ is (N, A) and N is a near-litter, where $\beta = \mathtt{min}(A_1)$, and $N^\circ = f_{\gamma,\beta}(x, T)$ with $-1 \leq \gamma < \mathtt{min}(A_2)$ and $x \in \tau_\gamma$ then S^*_ϵ is $(2, \beta, \chi_{x,T}, F, A)$, where F is a function from the domain of T into ϵ such that $S_{F(\delta)} = (T^{\uparrow A_2})_\delta$ for each δ in the domain of T, or 1 if there is no such F (the usefulness of 1 as a dummy being that it is not a function). There is a method to choose a canonical such F if there is one: add the provision that for each δ, $F(\delta)$ is chosen as small as possible.

Remark 4.27. It should be evident that every support has a specification, and that a strong support will have a specification with no instances of $F = 1$, and that for any β-allowable permutation π and strong β-support S, $(\pi[S])^* = S^*$. What is less evident and our first target result here is that if S is a strong support then any T with $T^* = S^*$ is the image of S under the action of an allowable permutation: the specifications precisely code the orbits in the strong supports under the allowable permutations.

Proposition 4.28. Suppose that we already know that there are $< \mu$ γ-coding functions for each $\gamma < \beta$ (which we will be able to assume by inductive hypothesis). Then there are $< \mu$ specifications of β-supports for $\beta \leq \alpha$.

Proof. The elements of the range of a β-specification are taken from a set of size less than μ, since they are built from ingredients in κ (support domain elements), $\beta + 1$ (type indices currently in use), and γ-coding functions for $\gamma < \beta$; each

of the collections from which the ingredients are taken are of size $< \mu$ and μ has cofinality at least κ. Therefore, the cardinality of the set of specifications is bounded by ν^ξ where $\nu < \mu$ and $\xi < \kappa$. Now, ν^ξ is the cardinality of the set of functions from ξ to ν, which is less than or equal to the cardinality of the power set of $\nu \times \xi$, which in turn is less than μ because μ is strong limit. □

A detail: We need to count *all* coding functions, allowing $\gamma < \beta$ to vary. We have $\beta < \lambda \leq \mathtt{cf}(\mu)$ and so the sum of cardinals each $< \mu$ indexed by ordinals $< \beta$ will be less than μ.

Lemma 4.29. The specification(s) of a strong β-support exactly determine the orbit in the action of β-allowable permutations on supports to which it belongs: if two β-supports have the same specification, they are in the same orbit.

Proof. It is straightforward to see that if S is a β-support and if π is a β-allowable permutation, and S^* is the specification for S, that S^* is also the specification for $\pi[S]$. 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 $S^* = T^*$ is a specification for both, there is an allowable permutation π such that $\pi[S] = T$.

We construct π using the Freedom of Action Theorem (4.18). Define the β-approximation ψ in the following way.

If $S_\epsilon = (M, A)$ for M a near-litter, then $T_\epsilon = (N, A)$ for N a near-litter, and we set $\psi_A(M^\circ) = N^\circ$. Note that if $S_\delta = (M, A)$ for $\delta \neq \epsilon$, the fact that S and T have the same specification ensures that $T_\delta = (N, A)$.

If we have $S_\epsilon = (\{x\}, A)$, we will have $T_\epsilon = (\{y\}, A)$ for some y, and we will set $\psi_A(x) = y$. Again, if $S_\delta = (\{x\}, A)$, then $T_\delta = (\{y\}, A)$.

We then complete orbits of atoms, with the proviso that if $x \in M$ where $S_\delta = (M, A)$, then $\psi_A(x) \in N$ where $T_\delta = (N, A)$. During this process, we ensure that for every near-litter M with $(M, A) \in \mathtt{rng}(S)$, we have $M \Delta M^\circ \subseteq \mathtt{dom}(\psi_A)$, and similarly, whenever $(N, A) \in \mathtt{rng}(S)$, we have $N \Delta N^\circ \subseteq \mathtt{dom}(\psi_A)^{-1}$. This process makes ψ into an approximation, and ensures that $\psi_A^*(M) = N$.

We now need to check that ψ is coherent. If M is A-flexible and we assigned $\psi_A(M^\circ) = N^\circ$, then $(M, A) = S_\epsilon$ and $(N, A) = T_\epsilon$ for some ϵ. As S and T have the same specification, M is A-flexible if and only if N is A-flexible.

Now suppose that $M^\circ = f_{\delta,\gamma}(x, S')$ where $\gamma = \mathtt{min}(A_1)$ and $\delta < \mathtt{min}(A_2)$. Again, as S and T have the same specification, we have

1. $N^\circ = f_{\delta,\gamma}(y, T')$;
2. there is a function F such that $S_{F(\delta)} = ({S'}^{\uparrow A_2})_\delta$ and $T_{F(\delta)} = ({T'}^{\uparrow A_2})_\delta$; and
3. $\chi_{x,S'} = \chi_{y,T'}$.

As $\chi_{x,S'} = \chi_{y,T'}$, there is a δ-allowable permutation π such that $\pi[S'] = T'$ and $\pi(x) = y$. By the second property and the definition of ψ, we have $(\psi_{A_2})^*_\delta[S'] = T' = \pi[S']$. It remains to check that

$$\psi_A^+ \text{“} f_{\delta,\gamma}(x, S') \sim f_{\delta,\gamma}(\pi(x), \pi'[S])$$

but this follows directly from the fact that $M^\circ = f_{\delta,\gamma}(x, S')$ and $N^\circ = f_{\delta,\gamma}(\pi(x), \pi'[S])$.

Finally, by the Freedom of Action Theorem (4.18), ψ approximates some β-allowable permutation π, and in particular, π satisfies $\pi[S] = T$, as required. □

Proposition 4.30. Under the inductive hypothesis that for each $\beta < \alpha$ we have $< \mu$ β-coding functions, there are less than μ orbits in supports under β-allowable permutations.

Proof. Since the β-specifications $[\beta \leq \alpha]$ precisely determine the orbits in strong supports under β-allowable permutations, and there are $< \mu$ β-specifications (on stated hypotheses) there are $< \mu$ such orbits.

Notice that we can give a kind of specification for any support S: give the specification for a strong support T of which S is an end extension and the index at which S starts: this will determine the orbit in which S lies in the allowable permutations. This establishes that the collection of orbits in the β-supports is no larger than the collection of orbits in the strong β-supports. These weak specifications are not unique. □

The strategy of our argument for the size of the types is to show that that there are $< \mu$ coding functions 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).

Lemma 4.31. There are less than μ coding functions for type 0 and for type -1

Proof. We describe all coding functions for type 0. The orbit of a 0-support in the allowable permutations is determined by the positions in the support occupied by near-litters, and for each position in the support occupied by a singleton, the positions, if any, of the near-litters in the support which include 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. 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.

Any type -1 coding function is associated with a uniquely determined type 0 coding function of a singleton, so there are no more of the former than of the latter.

□

Lemma 4.32. Under the inductive hypothesis that for each $\beta < \alpha$ we have $< \mu$ β-coding functions, there are less than μ coding functions for α.

Proof. By lemma 4.31 we may assume α is positive. Note that we already know that there are $< \mu$ α-specifications, on the stated inductive hypothesis (proposition 4.28).

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. ($\beta < \alpha$ is chosen arbitrarily here, and could for example be chosen to be 0).

We define S_b as the designated strong support for b, and T_b as the canonical extension to a strong support of $S_b^{\uparrow\alpha} + S$. *(In Lean we just use $S_b^{\uparrow\alpha} + S$ directly, or rather, $S + S_b^{\uparrow\alpha}$, because we don't need to worry about the terminal segment property of strong support extensions.)*

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} [we know that there are $< \mu$ β-coding functions] followed by a specification for T_b [we know that there are $< \mu$ α-specifications], 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. *(Sky: In my opinion this result is a good lemma. I'll think about it and get back to you. Holmes: leaving this comment in to encourage consderation of the revision)*

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 $\chi_{\{b\}_\alpha,T_b}$ for $b \in B$. There are $< \mu$ coding functions $\chi_{\{b\}_\alpha,T_b}$ for $b \in \tau_\beta$, and so there are $< \mu$ sets of coding functions of this kind, because μ is strong limit, and we have shown above in proposition 4.30 that there are $< \mu$ orbits in the strong α-supports under allowable permutations, so this will imply that there are $< \mu$ α-coding functions, which will further imply that there are $\leq \mu$ elements of type α (it is obvious that there are $\geq \mu$ elements of each type).

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 a terminal segment of U'. Clearly this definition depends only on the orbit of S and the set of coding functions $\chi_{\{b\}_\alpha,T_b}$ 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 $\chi^*_{X,S}(\pi[U]) = \pi(\chi^*_{X,S}(U))$. 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 has S as a terminal segment.

An arbitrary $c \in \chi^*_{X,S}(S)$ is of the form $\bigcup(\chi_{\{b\}_\alpha,T_b}(U) \cap \tau_\beta)$, where U has S as a terminal segment 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

$$c = \bigcup(\chi_{\{b\}_\alpha,T_b}(U) \cap \tau_\beta) = \bigcup(\pi_0(\chi_{\{b\}_\alpha,T_b}(T_b)) \cap \tau_\beta) = \bigcup(\pi_0(\{b\}_\alpha) \cap \tau_\beta)$$

$$= \bigcup(\{(\pi_0)_\beta(b)\}_\alpha \cap \tau_\beta) = (\pi_0)_\beta(b).$$

□

Thus, we conclude

Theorem 4.33. Each type τ_α has exactly μ elements.

Proof. Any element of a type is determined by a support (of which there are μ by Remark 3.25) and a coding function (there are $< \mu$ of these by lemma 4.32), so a type has no more than μ elements (and obviously has at least μ elements). □

4.3 The structure is a model of predicative TTT

There is then a very direct proof of the following:

Proposition 4.34. 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).

Proof. Use E for the membership relation $\in_{TTT}$ of the structure defined above (in which the membership of type β objects in type α objects is actually a subrelation of the membership relation of the metatheory, a fact inherited from the scheme of supertypes). 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 $\{x : \phi^s\}_{s(j+1)}$ 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), but it also works for some impredicative set definitions. □

There are easier proofs of the consistency of predicative tangled type theory; *[I'd be interested to know more!]* 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.

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 the following result.

Proposition 4.35. 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).

Proof. Suppose that $\{\{g\}_\beta : g \in G\}_\alpha$ is symmetric. It then has a strong support S. We claim that $S_{(\beta)}$, defined as $\{(z, C) : \mathtt{max}(C) = \beta \wedge (z, C \cup \{\alpha\}) \in S\}$, 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 a β-allowable permutation π fixes $S_{(\beta)}$.

Our plan is to use Freedom of Action technology to construct an α-allowable permutation π^* whose action on S is the identity and whose action on $T^{\uparrow\{\alpha,\beta\}}$ precisely parallels the action of π on $T^{\uparrow\beta}$.

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_β'.

We construct the allowable permutation π^* by the Freedom of Action Theorem (4.18), by defining the following α-approximation ψ.

For every $(\{x\}, A) \in \mathtt{rng}(S)$, we define $\psi_A(x) = x$, and for every $(N, A) \in \mathtt{rng}(S)$ for N a near-litter, we define $\psi_A(N^\circ) = N^\circ$. Extend $T^{\uparrow\beta}$ its canonical strong β-support T^* (remark 4.23), then if $(\{x\}, A) \in \mathtt{rng}(T^*)$, we define $(\psi_\beta)_A(\pi^n_A(x)) = \pi^{n+1}_A(x)$ for every integer n, and if $(N, A) \in \mathtt{rng}(T^*)$ for N a near-litter, we similarly define $(\psi_\beta)_A(\pi^n_A(N)^\circ) = \pi^{n+1}_A(N)^\circ$.

We then complete orbits of atoms similarly to the proof of lemma 4.29 to ensure that $\psi_A(N) = N$ for $(N, A) \in \mathtt{rng}(S)$ and $(\psi_\beta)_A(\pi^n_A(N)) = \pi^{n+1}_A(N)$ for $(N, A) \in \mathtt{rng}(T^*)$, and then arrange extra orbits if needed such that the interference condition in definition 4.1 is satisfied. *[Sky: Are extra details needed?]*

It now suffices to check that this approximation is coherent. If ψ^*_A"$L \sim L'$, we have two cases: either (N, A) appears in S with $N^\circ = L = L'$, or $B = A \setminus \{\alpha\}$ has maximum element β and (N, B) occurs in T^* with $\pi^n_B(N)^\circ = L$ and $\pi^{n+1}_B(N)^\circ = L'$. Note that if both of these cases are true at the same time, then as π fixes $S_{(\beta)}$, the two defined images of L coincide.

In the first case, if L is A-flexible then $L' = L$ is A-flexible as required, and if $L = f_{\delta,\mathtt{min}(A_1)}(x,T)$ with $\delta < \mathtt{min}(A_2)$, then as S is a strong support, the range of $T^{\uparrow A_2 \cup \{\delta\}}$ is a subset of the range of S, as required.

For the second case, we first show that L is A-flexible if and only if L is B-flexible. The forward direction is trivial. The converse can only fail if B is too short to satisfy the definition of inflexibility, so we may suppose that $L = f_{\delta,\mathtt{min}(A_1)}(x,T)$ with $\delta < \mathtt{min}(A_2)$ where A has length exactly 3. Then we must have $\mathtt{min}(A_1) = \beta$, so $A = \{\alpha, \beta, -1\}$ and $B = \{\beta, -1\}$. As (N, B) occurs in T^*, it either came from a support condition in T, which is impossible as B does not contain γ, or it came from a type-raised copy of an η-support we needed to include in T^* to make it a strong support. But $\eta \neq -1$ as -1-supports cannot contain near-litter support conditions, and if η were a proper type index, we would have $-1 < \eta < \beta$ and $\eta \in B$, also giving a contradiction. Hence $(T^*)^{\uparrow\alpha}$ is a strong α-support, and we are done by the method used for S.

□

This completes the proof. In the formal proof in Lean, what is actually done is a proof that each of the assertions in the finite axiomatization of Hailperin in the version discussed in subsection 2.5 holds in all typed versions in our structure for the language of TTT, so it is in fact a model of TTT. The axioms in Hailperin other than the axiom of type lowering are predicative comprehension axioms and admit demonstration by the methods of proposition 4.34 [or] section 4.3, done explicitly without metamathematics. In the formalization, the axiom of type lowering, which contains rather more content than the axiom of set union restricted to sets of singletons which is proved here, is proved by first proving the existence of an iterated image under elementwise application of the singleton operation of the desired set, whose definition is predicative, then repeatedly applying the result of this section that sets of singletons have unions.

5 Conclusions, extended results, and questions

In a series of subsections, we discuss further results supported by our model construction and further questions.

5.1 Ways in which this resolution of the NF consistency problem is unexciting

This is in some ways 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.

5.2 Questions and a result about global choice-like statements

Questions of consistency with NF of global choice-like statements such as "the universe is linearly ordered" or the prime ideal theorem cannot be resolved by the method used here (at least, not without major changes). Our models falsify the existence of a global linear order at the outset: a litter cannot be well-ordered. We see no reason to believe that the universe cannot be linearly ordered in NF or that the prime ideal theorem cannot hold, but we do not know how to model these situations.

We do have one quite interesting theorem which appears to show that most practical applications of the axiom of choice are supported in our models of TTT and so with care in the resulting models of NF.

Note that in our models of TTT, any relation which the internal theory thinks is a well-ordering actually is a well-ordering. The reason is that in the metatheory any linear order which is not a well-ordering admits a countable set with no minimal element (the range of a strictly decreasing sequence). This requires choice, but choice holds in the metatheory. So, if a relation represented in the model of TTT is not a well-ordering from the standpoint of the metatheory, there is a countable subcollection of its domain which has no minimal element, and this countable subset is actually realized in the model of TTT, because our model of TTT are countably complete, so the model of TTT sees that the relation is not a well-ordering.

Eassentially the same argument shows that any relation which a model of

TTT thinks is well-founded is actually well-founded in the sense of the metatheory.

Theorem: In our models of TTT, the power set of any well-orderable set can be well-ordered.

Proof: Suppose that X is a set in a model of TTT constructed by our method which admits a well-ordering (defined along a particular type sequence). The well-ordering $\leq_X$ of X which we are given must have a support. A permutation which fixes the support of $\leq_X$ [in the appropriate sense defined in the paper] will fix $\leq_X$ and, because $\leq_X$ is a well-ordering in the sense of the metatheory, its action on the type of the elements of X (in the type sequence with respect to which $\leq_X$ is a well-ordering) must fix each element of X, so in fact we have a common support for all elements of X, which in turn gives a support for each subcollection of X in the sense of the metatheory as implemented in the appropriate type in the particular type sequence, so all of these are sets in the model of TTT (in the appropriate type in the particular type sequence) and further gives a support for the well-ordering of the power set of X induced by any well-ordering of the power set of X in the sense of the metatheory, so the power set of X is well-ordered in the model of TTT.

This is likely reasonably easy to formalize, Sky

It is an interesting incidental remark that the theorem is equivalent in NF to the assertion that a single big power set, the power set of the set of ordinals, can be well-ordered, and this is seen to hold in models of NF constructed from our models of TTT.

This form of choice seems to support all practical applications of choice in mathematics outside of set theory, and a lot of the applications inside set theory.

Applying the same technique to sets with well-founded extensional relations on them can be used to show that all $\beth$ numbers are alephs, and that the internal representation of an initial segment of the cumulative hierarchy using equivalence classes of well-founded extensional relations with top which is possible in TTT will agree precisely with the cumulative hierarchy in the metatheory as far as it goes (the internal representation will always give a proper initial segment of the hierarchy in the metatheory, of course).

5.3 Strong unstratified axioms for NF derived from strong partition properties

NF with strong axioms such as the Axiom of Counting (introduced by Rosser in [14], an admirable textbook based on NF), the Axiom of Cantorian Sets (introduced in [5])[10] or my axioms of Small Ordinals and Large Ordinals (introduced

[10]Getting Cantorian Sets or Large Ordinals to hold is very sensitive to the relationship between κ and λ, and the first author is not yet entirely certain of the details. It requires the hypothesis that $\kappa < \lambda = \mu$ to avoid outright refutability of this axiom in resulting models

in my [7], 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 [8] on strong axioms of infinity in NFU: the results in that paper are not all mine, and I owe a good deal to Robert Solovay in that connection (unpublished conversations and [18]).

That NF has α-models for each standard ordinal α follows by the same methods Jensen used for NFU in his original paper [10]: the proof we give above for the existence of α-models is essentially the same as Jensen's, and a model of TTT constructed by our method will implement each ordinal α of the metatheory which is less than κ (and we can choose κ larger than any fixed α and λ with cofinality greater than $|\alpha|^{|\alpha|}$).

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.

5.4 Some esoteric questions answered

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 [22]) has ω-models; an ω-model of NF gives an ω-model of the theory of negative types immediately. The question of existence of ω models of the theory of negative types 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 methods show the existence of a model of NF with the Axiom of Counting. This gives us a model of TST in which the Specker tree of some cardinal is of infinite rank. It is straightforward to produce a model of bounded Zermelo set theory either with atoms or with extensionality and failure of foundation from a model of TST, which will satisfy the same stratified sentences: identify the elements of type 0 either as atoms or with their own singletons, then implement elements of each type $i + 1$ as subsets of

of NF, and then a large cardinal hypothesis. We note that for the moment we need to use Henson's original formulation of the Axiom of Cantorian Sets, that any *well-orderable* cantorian set is strongly cantorian; the situation for non-well-orderable sets is less clear.

type i as guided by the original model of TST, identifying type $i+1$ objects with elements of lower types which have been assigned the same extension. This shows that we can get a model of bounded Zermelo set theory with atoms or without foundation in which there is a cardinal with Specker tree of infinite rank. It is straightforward with our methods to get a model of TTT with the Axiom of Counting with inaccessibles cofinal in λ, and in this model we will have an inaccessible greater than a cardinal with Specker tree of infinite rank, and so a model of ZFA in which there is a cardinal with Specker tree of infinite rank. It's actually a bit easier than this: the existence of a cardinal with Specker tree of infinite rank holds in any model built by our method in which $\lambda > \omega$, but referring back to the result about NF + Counting makes it easier to give a brief account here.

The further question remains as to whether it is consistent with ZF that there is a cardinal with Specker tree of infinite rank. We would guess that this could be shown to be true using symmetric forcing extensions, but we are not expert at this and leave the question to others.

5.5 The question of consistency strength

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 ([9]). 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.

5.6 The wide open question: what do models of NF look like in general?

Another question of a very general and amorphous nature which remains is: what do models of NF (or TTT) 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.

5.7 Postscript

Inevitably, philosophical issues come up in connection with a system of set theory proposed by a philosopher. I get a lot of congratulations for vindicating Quine's foundational agenda, but in fact this is not part of my purpose here. It is not even clear to me that Quine *had* a foundational agenda in which his technical proposal of this set theory had a special place.

The results of this paper show that if one really wanted to, one could use NF as a foundation for mathematics. It is odd that it disproves Choice, but the

principle "power sets of well-orderable sets are well-orderable", which holds in models constructed in this way, supports most applications of Choice.

Our opinion is that Quine's proposal of NF was based on a mistake. He discusses whether to assume strong extensionality in the original paper, and his explanation of reasons for choosing strong extensionality contains an actual mathematical error. *I would be very interested in details.* We believe that the correct system to propose was NFU, and if he had proposed NFU, the history of this kind of set theory might have been different.

NFU is serviceable as a foundation for mathematics, and consistent with Choice and various strong axioms of infinity. It is odd that NFU + Choice proves that there are urelements (the same odd fact that NF disproves Choice), but no more than odd. The type discipline of stratification is something that one must work to get used to, and it has been remarked that working with indexed families of sets is extremely awkward in NFU (and would be similarly awkward in NF). The view of the world which NFU supports is basically the same as that of ZFC: the natural models of NFU are obtained by considering an initial segment of the cumulative hierarchy with an external automorphism which moves a level, and NFU can interpret discussion of exactly such a structure internally.

Unlike NF, NFU on introspection can tell one quite a lot about what its models should look like (as ZFC can with its own awareness of initial segments of the cumulative hierarchy; in NFU, analysis of the isomorphism classes of well-founded extensional relations gives both an interpretation of an initial segment of the Zermelo style universe and an interpretation of NFU itself if one has strong enough assumptions). NF tells one very little about what its intended world is like (it can, being an extension of NFU, internally construct a lot of information about an interpretation of NFU with lots of urelements, but there is not an obvious way to find out how an extensional world is constructed from internal evidence in NF itself).

So, we do not believe this paper is a philosophical milestone. If there was one, it happened in 1969 when Jensen showed that NFU is consistent, and nobody noticed.

We do believe that there are interesting questions to investigate about NF. Our paper does not settle all the questions about this system which have developed in the minds of people who have worked with it since 1937. In fact, the construction is based on special assumptions and does not seem to give much of an idea of what general models of this theory might look like.

It might be viewed as philosophically interesting that this proof was formally verified. That is really an advertisement for a quite different foundational system, the logic of the Lean proof verification system. That it needed to be formally verified I believe reflects an interesting complexity in the mathematics here, not only the deficiencies of the first author as an expositor.

References

[1] Crabbé, M. [1992a] On NFU. *Notre Dame Journal of Formal Logic* 33, pp 112-119.

[2] Scott Fenton. New Foundations set theory developed in metamath. 2015. `https://us.metamath.org/nfeuni/mmnf.html`

[3] 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.

[4] Hailperin, T. [1944] A set of axioms for logic. *Journal of Symbolic Logic* 9, pp. 1-19.

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

[6] 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.

[7] 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.

[8] 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).

[9] Kemeny, J.G. [1950] Type theory vs. set theory (abstract of Ph.D. thesis). *Journal of Symbolic Logic* 15, p. 78.

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

[11] Quine, W.V. [1945] On ordered pairs. **Journal of Symbolic Logic** 10, pp. 95-96.

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

[13] Quine, W.V. *Mathematical Logic*. Norton, 1940. revised ed. Harvard, 1951.

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

[15] Russell, Bertrand. *Principles of Mathematics*. Routledge Classics 2010 (originally published 1903).

[16] Russell, B.A.W. and Whitehead, A. N.[1910] *Principia Mathematica.* Cambridge University Press.

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

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

[19] 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.

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

[21] Tarski, Alfred. Einige Betrachtungen uber die Begriffe der co-Widerspruchsfreiheit und der co-Vollstandigkeit, `Monatsh. Math. Phys.` 40 (1933), 97-112.

I believe this is the right paper. There is a treatment due to Gödel at about the same time. The reason that I think Tarski is the originator of TST is that Gödel's theory is ω-order arithmetic (the lowest type is further equipped with the Peano axioms) which is not the same theory, and crucially does not have the ambiguity properties which motivate NF.

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

[23] Wiener, Norbert, paper on Wiener pair

[24] Sky Wilshaw, Yaël Dillies, Peter LeFanu Lumsdaine, et al. New Foundations is consistent. GitHub repository. `https://leanprover-community.github.io/con-nf/`.

Initial work was done by a group, as indicated by the authorship; since the beginning of 2023, the work, which now amounts to the majority of what has been accomplished, has been done by Sky Wilshaw alone, and the previous work has been very largely reorganized by her.