# The proof of consistency of New Foundations from a tangled web

M. Randall Holmes

finished transferring BEST slides, 6/30/2021
editing 7/1/2021
rough draft of the entire proof 7/9/2021
more edits later July 9
editing July 10

## 0.1   Version notes

**July 10, 2021** More editing.

**July 9, 2021:** Reached the end of the proof in rough form. Some additional editing.

**July 1, 2021:** Considerable editing of material from the slides. I still have to fix up the last subsection.

**June 30, 2021:** I've been working steadily. Finished transferring the BEST slides: it remains to work through these and see what clarifications are needed and to finish the argument for external size of double power sets of clans.

**June 28, 2021:** Starting.

# Contents

# 1 Development of relevant theories

## 1.1 The simple theory of types TST and TSTU

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

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

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

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 $\phi$ is any formula in which $A$ does not occur free.

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

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

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

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

### 1.1.1   Typical ambiguity

TST exhibits a symmetry which is important in the sequel.

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

If $\phi$ is a formula, define $\phi^+$ as the result of replacing every variable $x$ (free and bound) in $\phi$ with $x^+$. It should be evident that if $\phi$ is well-formed, so is $\phi^+$, and that if $\phi$ is a theorem, so is $\phi^+$ (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 $\phi$ is called the Ambiguity Scheme. Notice that this is a stronger assertion than is warranted by the symmetry of proofs described above.

### 1.1.2   Historical remarks

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

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

It was shown around 1950 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.

## 1.2 Some mathematics in TST; the theories $\text{TST}_n$ and their natural models

We briefly discuss some mathematics in TST.

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

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

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

The Kuratowski pair $(x, y) = \{\{x\}, \{x, y\}\}$ is of course two types higher than its projections, which must be of the same type. There is an alternative definition (due to Quine) 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 $\iota$ 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). 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

7

usual definition $|A| \cdot |B| = |A \times B|$. Use of the Quine pair simplifies matters in this case, but it should be noted that the T operation remains quite important (for example it provides the internally representable isomorphism between the systems of natural numbers in each sufficiently high type).

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

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

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

Notice then that we can define truth for formulas in these natural models of $\mathrm{TST}_n$ for each $n$ in TST, though not in a uniform way which would allow us to define truth for formulas in TST in TST.

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

## 1.3 New Foundations and NFU

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

Quine's observation is that since any theorem $\phi$ 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 $\phi$ 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 $\phi$ is a formula of the language of TST, let $\phi^*$ be the formula obtained by replacing every veriable $x$, free and bound, in $\phi$ with $x^*$. For each formula $\phi$ of the language of TST, an axiom of comprehension of NF asserts $(\exists A : (\forall x : x \in A \leftrightarrow \phi^*))$.

In the original paper, this is expressed in a way which avoids explicit dependence on the language of another theory. Let $\phi$ be a formula of the language of NF. A function $\sigma$ is a stratification of $\phi$ if it is a (possibly partial) map from variables to non-negative integers such that for each atomic subformula '$x = y$' of $\phi$ we have $\sigma('x') = \sigma('y')$ and for each atomic subformula '$x \in y$' of $\phi$ we have $\sigma('x') + 1 = \sigma('y')$. A formula $\phi$ is said to be stratified iff there is a stratification of $\phi$. Then for each stratified formula $\phi$ of the language of NF we have an axiom $(\exists A : (\forall x : x \in A \leftrightarrow \phi))$. The stratified formulas are exactly the formulas $\phi^*$ 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, refined by others. One obtains a finite axiomatization of NF by analogy with the method of finitely axiomating von Neumann-Gödel-Bernays predicate class theory. It should

further be noted that the first thing one does with the finite axiomatization is prove stratified comprehension as a meta-theorem, in practice, but it remains significant that the theory can be axiomatized with no reference to types at all.

For each stratified formula $\phi$, 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 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 $\phi$ 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 (1962) is that NF is consistent if and only if TST + the Ambiguity Scheme is consistent. His method of proof adapts to show that NFU is consistent if and only if TSTU + the Ambiguity Scheme is consistent. Jensen used this fact in his proof of the consistency of NFU. We indicate a proof of Specker's result using concepts from this paper below.

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

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

The apparent vulnerability to the paradox of Cantor is an illusion. Applying Cantor's theorem to the cardinality of the universe in NFU gives $|\iota "V| < |(V)| \leq |V|$ (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.

## 1.4 Tangled type theory TTT and TTTU

In 1995, this author described a reduction of the NF consistency problem to consistency of a typed theory and of a kind of extension of bounded Zermelo theory, both motivated by reverse engineering from Jensen's method of proving the consistency of NFU.

Let $\lambda$ be a limit ordinal. It can be $\omega$ 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 $\mathtt{type}(`x\text') < \lambda$; we are provided with countably many distinct variables of each type.

For any formula $\phi$ of the language of TST and any strictly increasing sequence $s$ in $\lambda$, let $\phi^s$ 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 $\alpha$ 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 $\lambda$ whose well formed atomic sentences '$x = y$' have $\mathtt{type}(`x\text') = \mathtt{type}(`y\text')$ and whose atomic sentences '$x \in y$' satisfy $\mathtt{type}(`x\text') < \mathtt{type}(`y\text')$, and whose axioms are the sentences $\phi^s$ for each axiom $\phi$ of TST and each strictly increasing sequence $s$ in $\lambda$. TTTU has the same relation to TSTU (with the addition of constants $\emptyset^{\alpha,\beta}$ for each $\alpha < \beta < \lambda$ such that $(\forall \mathbf{x}_0^\alpha : \mathbf{x}_0^\alpha \notin \emptyset^{\alpha,\beta})$ is an axiom).

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

There is now a striking

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

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

Suppose TTT(U) is consistent, and so we can assume we are working with a fixed model of TTT(U). Let $\Sigma$ be a finite set of sentences in

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

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

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

**Proof:** It is enough to exhibit a model of TTTU. Suppose $\lambda > \omega$. Represent

13

type $\alpha$ 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 $\alpha$ in type $\beta$ interprets each $(y,\beta)$ 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 $\alpha$ using $V_{\alpha}$ 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 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.

### 1.4.1 How internal type representations unfold in TTT

We have seen above that TST can internally represent $\mathrm{TST}_n$. An attempt to represent types of TTT internally to TTT has stranger results.

In TST the strategy for representing type $i$ in type $n \geq i$ is to use the $n-i$-iterated singleton of any type $i$ object $x$ to represent $x$; then membership of representations of type $i - 1$ objects in type $i$ objects is represented by the relation on $n - i$-iterated singletons induced by the subset relation and

with domain restricted to $n - (i+1)$-fold singletons. This is described more formally above.

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

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

the natural numbers of our model of TTT are standard.

## 1.5 Tangled webs defined

The situation just described in a model of TTT can be transferred to ZF or fragments thereof. We favor bounded Zermelo set theory with extensionality weakened to allow a set of atoms with the assumption that everything is in a rank of the cumulative hierarchy (the set of atoms being rank 0): this is no stronger than TST + Infinity and supports the Scott definition of cardinal (we do not want to assume choice, because the situation we are about to described refutes choice). Recall that the Scott cardinal of a set $A$ is the set of all sets $B$ such that $B \sim A$ which are of minimal rank.

A tangled web is a function $\tau$ from nonempty finite subsets of a limit cardinal $\lambda$ to cardinals with the following properties

**naturality:** For each $A$ with $|A| \geq 3$, $2^{\tau(A)} = \tau(A_1)$

**elementarity:** For each $A, B$ with $A \backslash A_{n+1} = B \backslash B_{n+1}$ and $n+1 < |A| \leq |B|$, the theories of natural models of $\text{TST}_n$ with base types $\tau(A)$ and $\tau(B)$ will be the same.

The conditions stated here are weaker than those which hold in the implementation $\tau(A) = |\tau_A|$ in TTT suggested by the discussion above (one might reasonably expect 2 instead of 3 in naturality and $n + 1$ instead of $n$ in elementarity); the weaker conditions are what we can conveniently make hold in our construction below, and they turn out to be sufficient for us to show the following

**Theorem:** The existence of a tangled web implies the consistency of NF

**Proof:** Suppose that there is a tangled web $\tau$. Let $\Sigma$ be a finite set of formulas in the language of TST. Let $n$ be any natural number dominating all types which occur in a formula in $\phi$. We define a partition of $[\lambda]^{n+1}$ (where $\tau$ is defined on nonempty subsets of $\lambda$), putting each $A \in [\lambda]^{n+1}$ in a partition determined by the truth values of the formulas in $\Sigma$ in natural models of TST with base type of cardinality $\tau(B)$ with $|B| > n+1$ and $B \setminus B_{n+1} = A$. This partition has a homogeneous set $H$ of size $n + 2$. A natural model of $\text{TST}_{n+1}$ with base type $\tau(H)$ satisfies $\phi \leftrightarrow \phi^+$ for each $\phi \in \Sigma$, because $\phi$ is satisfied in this model iff $\phi^+$ is satisfied in the model of $\text{TST}_n$ whose base type is type 1 of this natural model, which has size $2^{\tau(H)} = \tau(H_1)$, which has the same

17

theory by homogeneity of $H$ with respect to the partition. Consistency of TST + Ambiguity follows by compactness, since we have shown that $\text{TST}_{n+1}$ is consistent with ambiguity for any finite set of formulas of the language of TST for any large enough $n$.

The actual existence of a tangled web in a model of TTT would follow if the model had standard natural numbers and every subset of a type in the model with cardinality less than or equal to the cardinality of $\lambda$ was implemented in each higher type in the model.

It is also worth noting that the definition of a tangled web and the proof that existence of a tangled web implies consistency of NF do not depend on understanding anything about tangled type theory at all. Tangled type theory does have the advantage of motivating a neater argument for Specker's ambiguity result, which the tangled web approach doesn't support, but we can rely on Specker'original argument.

# 2  The construction of a tangled web in ZFA

We work in ZFCA (Zermelo-Fraenkel set theory with choice and with extensionality weakened to allow a set of atoms), with some assumptions about the size of the set of atoms which we will detail shortly. Consistency of this theory relative to ZFC is unproblematic. In this theory, we carry out a Fraenkel-Mostowski construction of an interpretation of ZFA in which there is a tangled web. This establishes the consistency of New Foundations.

## 2.1  The description of the postulated structure

In this section, we describe the structure which turns out to be the Fraenkel-Mostowski model containing a tangled web. It is rather elaborate.

### 2.1.1  Important cardinals and related concepts

We specify a limit cardinal $\lambda$. Our intention is to construct a tangled web with domain the nonempty finite subsets of $\lambda$. This ordinal can be $\omega$ but does not have to be.

If $A$ is a nonempty finite subset of $\lambda$, we define $A_1$ as $A \setminus \mathtt{min}(A)$. We define $A_{i+1}$ as $(A_i)_1$ if this is defined.

We let $\ll^*$ be the unique strict well-ordering of finite subsets of $\lambda$ satisfying these conditions ($A$, $B$ ranging over all finite subsets of $\lambda$):

1. $A \ll^* \emptyset$

2. If $\mathtt{max}(A) < \mathtt{max}(B)$, then $A \ll^* B$.

3. If $\mathtt{max}(A) = \mathtt{max}(B)$, then $A \ll^* B \leftrightarrow A \setminus \{\mathtt{max}(A)\} \ll^* B \setminus \{\mathtt{max}(B)\}$.

We further define $A \ll B$ as ($\exists i \in \mathbb{Z}^+ : A_i = B$): that is, $B$ is a proper initial segment of $A$ in the natural order. It should be clear that $A \ll B \to A \ll^* B$.

We specify an regular cardinal $\kappa > \lambda$ (and so uncountable). If $\lambda = \omega$, we can have $\kappa = \omega_1$; the generality of allowing larger values is important for our conclusions. We call sets of cardinality $< \kappa$ *small* and all other sets *large*. The intention is that all subsets of the model of ZFA with a tangled web which are small will be sets in the model of ZFA, so the meaning of "small" will be unequivocal (the same in the ambient set theory and in the FM interpretation).

We specify a strong limit cardinal $\mu > \kappa$ whose cofinality is at least $\kappa$. If $\lambda = \omega$, we can have $\kappa = \omega_1$ and $\mu = \beth_{\omega_1}$.

### 2.1.2　The structure of the set of atoms; clans and litters

There are $\mu$ atoms.

We provide a partition of the atoms into disjoint sets $\tau_A^0$ indexed by nonempty finite subsets of $\lambda$, each of these sets being of size $\mu$. These sets may be referred to as *clans*: the notation for $\tau_A^0$ in older accounts of this argument is `clan[A]`.

We provide a partition of each $\tau_A^0$ into sets of size $\kappa$ which we refer to as *litters*. A subset of a clan with small symmetric difference from a litter we call a *near-litter*. If $N$ is a near-litter, we denote the litter with small symmetric difference from $N$ by $N^\circ$. Elements of $N \Delta N^\circ$ we may refer to as anomalies or anomalous elements of $N$. The collection of all near-litters $N$ such that $N^\circ = L$ we call the local cardinal of $L$ and denote by $[L]$. We will see that in the FM model the local cardinal of $L$ will be the intersection of the Scott cardinal of $L$ with the clan to which $L$ belongs. We define $K_A$ as the set of local cardinals of litters included in $\tau_A^0$.

We define $\tau_A^1$ as the set of all subsets $X$ of $\tau_A^0$ for which there is a small set $Y$ of litters included in $\tau_A^0$ such that either $X \Delta \bigcup Y$ is small or $X \Delta (\tau_A^0 \setminus \bigcup Y)$ is small. In other words, $\tau_A^1$ is the collection of all subsets of $\tau_A^0$ which have small symmetric difference from small or co-smal unions of litters. We will see that the FM model sees $\tau_A^1$ as the power set of $\tau_A^0$.

We will construct subsets $\tau_A^2$ of each $\tau_A^1$ whose description is rather elaborate (recursive along $\ll^*$). These sets are each of size $\mu$. The FM model we eventually construct will see $\tau_A^2$ as the power set of $\tau_A^1$. Further, the value of the tangled web at $A$ in the FM model will be the FM model's cardinality for $\tau_A^2$.

### 2.1.3 Isomorphisms on the $\tau_A^i$'s; parent maps

For each $\alpha$ we choose a map $\chi_\alpha$ which is an injective map with domain the union of all $\tau_A^0$ with $\mathtt{max}(A) = \alpha$ and $|A| > 1$ whose restriction to each such $\tau_A^0$ is a bijection from $\tau_A^0$ to $\tau_{A\setminus\{\mathtt{max}(A)\}}^0$. We further require that the elementwise image of each litter under $\chi_\alpha$ is a litter.

We extend the action of $\chi_\alpha$ to any set whose transitive closure contains no atoms not in its domain by the rule $\chi_\alpha(X) = \chi_\alpha \text{``} X$.

The motivation for our use of maps $\chi_A$ is to provide a way to in effect identify elements of sets $\tau_A^i$ with elements of $\tau_{A\setminus\{\mathtt{max}(A)\}}^i$, via the map $\chi_{\mathtt{max}(A)}$: recall our intuitive motivation of tangled webs by "unfolding" types in TTT, in which all representations of types indexed by sets with the same minimal element are to be understood as secretly representing he same type.

We will provide for each pair $\{\alpha, \beta\}$, $\alpha > \beta$, a map $\Pi_{\{\alpha,\beta\}}$, a bijection from $K_{\{\alpha,\beta\}}$ to the union of $\tau_{\{\alpha\}}^0$ and all sets $\tau_{\{\alpha,\beta,\gamma\}}^2$ for which $\gamma < \beta$. For each $A$ with $|A| \geq 3$ we define $\Pi_A$ as $\chi_{\mathtt{max}(A)}^{-1}(\Pi_{A\setminus\{\mathtt{max}(A)\}})$.

Notice that when we define $\Pi_{\{\alpha,\beta\}}$, we will suppose that we have already defined all $\tau_B^2$ and $\Pi_B$ for $B \ll^* \{\alpha, \beta\}$, and in particular $\tau_{\{\alpha,\beta,\gamma\}}^2$ is supposed already defined. The definition of any $\tau_B^2$ will presuppose that $\Pi_C$ has been defined for $C = B$ and $C \ll^* B$, and that $\tau_C^2$ has been defined for $C \ll^* B$.

It follows that $\Pi_A$ is a bijection from $K_A$ to the union of $\tau_{A_1}^0$ and the union of all $\tau_B^2$ for which $B_1 = A$.

Traditional terminology from older versions of this argument which I will feel free to use: when $x$ is an atom in the clan $\tau_A^0$ belonging to the litter $L$, to refer to $\Pi_A([L])$ as the *parent* of $x$, or of $L$, or for that matter of $[L]$ or of any $N \in [L]$.

### 2.1.4 Allowable permutations and supports

We allow a permutation $\pi$ of the set of atoms to induce a permutation of the entire universe by the rule $\pi(X) = \pi\text{``}X$ for each set $X$.

An $A$-allowable permutation is a permutation of atoms whose action fixes each $\tau_B^0$, fixes each $K_B$ (so it always maps litters to near-litters in the same clan), and fixes $\Pi_B$ for $B \ll^* A$. An $\emptyset$-allowable permutation is simply called an allowable permutation.

A small well-ordering of atoms and near-litters is called a support. An object $X$ has $A$-support $S$ iff $S$ is a support and each $A$-allowable permutation $\pi$ such that $\pi(S) = S$ also satisfies $\pi(X) = X$. A strict $A$-support is one for which if its domain meets $\tau_B^i$, we have $B \ll A$ or $B = A$ (i.e. $B$ is a downward extension of $A$).

We note that if an object has an $A$-support it certainly has an $A$-support whose domain contains only atoms and litters (replacing any near-litter $N$ with the nearby litter $N°$ and its anomalous atoms, the elements of $N \Delta N°$) but it is important to allow general near-litters in supports so that $A$-allowable permutations send supports to supports.

An object $X$ has prestrong $A$-support $S$ iff $X$ has $A$-support $S$ and each near-litter in the domain of $S$ is a litter, and each atom in the domain of $S$ which is not in $\tau_A^0$ belongs to a litter in the domain of $S$ preceding it in $S$ and for each litter $N \in \mathtt{dom}(S)$ belonging to $\tau_B^1$ for $B \ll A$ , the segment of $S$ before $N$ includes a strict $C$-support of $\Pi_B([N])$ where $\Pi_B([N]) \in \tau_C^i$ for $i = 0, 2$ [this strict $C$-support including $\Pi_B([N])$ as an element if $\Pi_B([N])$ is an atom], unless $C = A$.

An object $X$ has strong $A$-support $S$ iff $X$ has strict prestrong $A_1$-support $S$ (so $S$ is actually an $A_1$-support), and each element of $S$ belongs to a $\tau_C^i$ with $C \ll A$, $C = A$ or $C = A_1$.

Note that a prestrong $A$-support satisfies the condition that for each litter $N \in \mathtt{dom}(S)$ belonging to $\tau_B^1$ for $B \ll A$ , the segment of $S$ before $N$ includes a **strong** $C$-support of $\Pi_B([N])$ where $\Pi_B([N]) \in \tau_C^i$ for $i = 0, 2$, unless $C = A$. The closure conditions on the prestrong $A$-support enforce stronger conditions on embedded $B$-supports for $B \ll A$,

### 2.1.5 The definition of $\tau_A^2$ (mod a refinement)

The collection $\tau_A^2$ consists exactly of those subsets of $\tau_A^1$ which have strong $A$-supports (these will turn out to be exactly those which have $A_1$-supports).

This actually completes the definition of the sets $\tau_A^2$, mod the choice of the maps $K_B$ for $B \ll^* A$ and $B = A$, and an annoying refinement described in the next subsection, as long as we can verify that $\tau_A^2$ is of size $\mu$ in the ambient set theory.

The role of the maps $\chi_\alpha$ is to provide an isomorphism between sets $\tau_A^2$ and $\tau_{A \cup B}^2$ with respect to set theoretical structure and relevant maps $\Pi_C$ (mapped to $\Pi_{C \cup B}$) when all elements of $B$ dominate all elements of $A$: $\chi_{\max(A)}$ witnesses an isomorphism between $\tau_A^2$ and $\tau_{A \setminus \{\max(A)\}}^2$, and iteration of this fact gives the stated result.

### 2.1.6 Refinement of the choice of the parent maps; analysis of supports

We describe an annoying refinement of the choice of the maps $\Pi_{\{\alpha,\beta\}}$ which seems to be necessary.

We provide a well-ordering $<^1_{\alpha,\beta}$ of the union of $\tau^0_{\{\alpha\}}$ and all sets $\tau^2_{\{\alpha,\beta,\gamma\}}$ with $\gamma < \beta < \alpha$. We provide a well-ordering $<^2_{\alpha,\beta}$ of $K_{\{\alpha,\beta\}}$. We stipulate that both orders are of order type $\mu$. We define orders $<^i_A$: $<^i_A$ for $|A| > 1$ is the image under $\chi^{-1}_{\max(A)}$ of $<^i_{A\setminus\{\max(A)\}}$ $(i = 1, 2)$.

We regiment the construction of $K_{\{\alpha,\beta\}}$. The idea is that when we apply $\Pi_{\{\alpha,\beta\}}$ to an element $[L]$ of $K_{\{\alpha,\beta\}}$, we want to obtain, if the ordinal is even, the $<^1_{\alpha,\beta}$-first element of $\tau^0_{\{\alpha\}}$ not already used as a value at a $<^2_{\alpha,\beta}$-earlier element of $K_{\{\alpha,\beta\}}$, and if the ordinal is odd, the $<^1_{\alpha,\beta}$-first element not already used as a value at a $<^2_{\alpha,\beta}$-earlier element of $K_{\{\alpha,\beta\}}$ in the appropriate well-ordering of a $\tau^2_{\{\alpha,\beta,\gamma\}}$ which has an $\{\alpha, \beta\}$-prestrong support $S$ such that any element of the domain of $S$ which is an element $M$ of $\tau^1_{\{\alpha,\beta\}}$ has had $\Pi_{\{\alpha,\beta\}}([M])$ already defined (that is, $[M] <^2_{\alpha,\beta} [L]$).

We observe that any strict $B$-support included in a segment in a prestrong $A$-support ($B \ll A$, and so if the support is nonempty, $B_i = A$ for some $i$) is included in a strong $B$-support included in the same segment, because the closure conditions of a prestrong $A$-support ensure that the segment already includes suitable $B_1$-supports of each element of $\tau^2_B$ in the strict $B$-support.

A consequence of this is that every element of any $\tau^2_A$ has an $A$-strong support with the further property that for for each $L \in \tau^1_A$ which is in $S$, there is a $A$-strong support for $\Pi_A([L])$ included in the segment preceding $L$ with the property that for each $M$ in this support belonging to $\tau^1_A$, we have $[M] <^2_A [L]$.

A further consequence is that any strict $A_1$-support can be extended to an $A$-strong support. This is done by adding supports of litters appearing in the support which satisfy the condition just stated before the litter in question. This process can be iterated through $\omega$ stages to obtain an ordered set, which will be a well-ordering because it is impossible to have an infinite regress in the process of adding items to the support: a litter needed for a strong support of an element of $\tau^1_B$ will either be in $\tau^1_B$ and earlier in the well-ordering $<^2_B$, or will be in a $\tau^2_C$ with $C \ll B$. Similarly, any $A$-support can be extended to an $A$-prestrong support.

### 2.1.7 The Fraenkel-Mostowski models; claims to be checked

There are actually models determined by each finite subset $A$ of $\lambda$.

We say that a set is $A$-symmetric if it has an $A$-support. There is a model of ZFA consisting of the atoms and hereditarily $A$-symmetric sets, by standard results about the Fraenkel-Mostowski permutation method.

We use the notation $\mathcal{P}_*(X)$ for the power set of $X$ in an FM model. We use the notation $|X|_*$ for the Scott cardinal of $X$ in an FM model.

The model we are ultimately concerned with is the model consisting of the atoms and the hereditarily $\emptyset$-symmetric sets.

We list claims that we need to verify.

1. The cardinality of $\tau_A^2$ is $\mu$. This must be verified for each $B \ll A$ for $\tau_A^2$ to even be defined.

2. $\mathcal{P}_*(\tau_A^0) = \tau_A^1$ for any $A$ in any of the FM models.

3. $\mathcal{P}_*(\tau_A^1) = \tau_A^2$ in the model consisting of atoms and $B$-symmetric sets if $B = A_1$ or $A_1 \ll^* B$.

4. $\tau(A) = |\tau_A^2|_*$ defines a tangled web in the model consisting of atoms and $\emptyset$-symmetric sets.

Facts about any of the FM models which are readily verified are that each atom and each near-litter belongs to each FM model and any small subset of the doman of an FM model belongs to the same FM model.

## 2.2   Verification of the properties of the purported model

In this section, we verify that the structure defined above actually contains the tangled web indicated.

### 2.2.1 Mod the indicated claims, $\tau$ is a tangled web.

We verify that (subject to outstanding claims) we can show that $\tau(A) = |\tau_A^2|_*$ defines a tangled web in the FM model determined by all allowable permutations.

We use the notation $|X|_* \leq |Y|_*$ for assertions that the cardinal of $X$ is less than the cardinal of $Y$ in the FM model. We use the notation $2^{|X|_*}$ for $|\mathcal{P}_*(X)|_*$: we are allowing a certain amount of overloading but there should always be a notational reminder that we are working in an FM model, if we are working in an FM model.

Obviously $|K_A|_* \leq |\tau_A^2|_*$, since elements of $K_A$ are elements of $\tau_A^2$. An element of $K_A$, the local cardinal of a litter, has the well-ordering on the singleton of that litter as a support. Further, in fact $2^{|K_A|_*} \leq |\tau_A^2|_*$, because subsets of $K_A$ are in one to one correspondence with their set unions, which are elements of $\tau_A^2$, because $K_A$ is a pairwise disjoint collection. Because of the existence of the map $\Pi_A$, we have $|\tau_{A_1}^0|_* \leq |K_A|_*$ and $|\tau_B^2|_* \leq |K_A|_*$ when $B_1 = A$.

We define $\exp(|X|_*) = 2^{|X|_*}$.

The inequalities above further give $\exp(|\tau_{A_1}^0|_*) \leq |\tau_A^2|_*$ and $\exp(|\tau_B^2|_*) \leq |\tau_A^2|_*$ when $B_1 = A$, so $\exp(|\tau_A^2|_*) \leq |\tau_{A_1}^2|_*$ when $|A| \geq 3$.

Further, we get $\exp^2(|\tau_{A_1}^0|_*) = |\tau_{A_1}^2|_* = |\tau(A_1)|_* \leq \exp(|\tau_A^2|_*) = \exp(\tau(A))$. and $\exp(|\tau_A^2|_*) = \exp(\tau(A)) \leq |\tau_{A_1}^2|_* = \tau(A_1)$ (where $|A| \geq 3$), so we have the naturality property of a tangled web for $\tau$.

The natural model of $\mathrm{TST}_n$ with base type $\tau_A^2$ is sent by the composition of $\chi_\alpha$'s determined by the elements of $A_{n+1}$ to the natural model of $\mathrm{TST}_n$ with base type $\tau_{A \setminus A_{n+1}}^2$, and the $\chi_\alpha$'s are external isomorphisms for all relevant structure, so the first order theory of these models is the same. For this to make sense of course we need $|A| > n + 1$. The reason for this is that the size of type $i < n$ in the first model is internally seen to be the same as that of $\tau_{A_i}^2$, and type $i$ in the second is internally seen to be the same size as $\tau_{(A \setminus A_{n+1})_i}^2 = \tau_{A_i \setminus A_{n+1}}^2$, and independently of the value of $i$ the same composition of $\chi_\alpha$'s serves as an external isomorphism. This verifies the elementarity property of $\tau$.

### 2.2.2 The Freedom of Action theorem

If $\pi$ is an allowable permutation, we say that an atom $x$ is an exception of $\pi$ if either $\pi(x) \notin \pi(L)^{\circ}$ or $\pi^{-1}(x) \notin \pi^{-1}(L)^{\circ}$, where $L$ is the litter containing $x$.

Define a local bijection as a map from atoms to atoms which is injective, has domain the same as its range, sends elements of a given $\tau_A^0$ to elements of the same $\tau_A^0$, and whose domain has small intersection with each litter (empty being a case of small).

The Freedom of Action theorem asserts that for any $A$, any local bijection $\pi_0$ can be extended to an $A$-allowable permutation $\pi$ with the property that each exception of $\pi$ is either fixed by $\pi$ or belongs to the domain of $\pi_0$.

We commence proving the Freedom of Action theorem. Fix a local bijection $\pi_0$ and a finite subset $A$ of $\lambda$.

Specify a well-ordering $<_L$ of type $\kappa$ of each litter $L$. For each co-small subset $L'$ of a litter $L$ and co-small $M'$ of a litter $M$ define $\pi_{L',M'}$ as the unique bijection from $L'$ to $M'$ such that $\pi_{L',M'}(x) <_M \pi_{L',M'}(y)$ iff $x <_L y$, for all $x, y \in L'$.

For any atom $x$, we compute $\pi(x)$ by a recursion along an $A$-strong support of $x$.

If $x$ is in the domain of $\pi_0$, $\pi(x) = \pi_0(x)$.

If $x$ is in a $\tau_B^0$ with $A \ll B$, and not in the domain of $\pi_0$, $\pi(x) = x$. Alternatively, $\pi$ could be made to agree with an arbitrary $\pi'$ extending $\pi_0$ and sending local cardinals to local cardinals at such values.

For the remaining cases, in which $x \in \tau_B^0 \setminus \mathtt{dom}(\pi_0)$ and $B \ll A$ or $B = A$, we first compute $\pi(\Pi_B([L]))$, where $L$ is the litter to which $x$ belongs, then $\pi(x) = \pi_{L \setminus \mathtt{dom}(\pi_0), \pi(L)^{\circ} \setminus \mathtt{dom}(\pi_0)}(x)$, where $\pi(L)^{\circ}$ is the litter in $\Pi_B^{-1}(\pi(\Pi_B([L])))$.

It should be evident that what we have said already enforces that $\pi$ has no exceptions outside the domain of $\pi_0$.

It remains to say how to compute $\pi(\Pi_B([L]))$.

We note that $L$ precedes $x$ in the strong support, and we assume as an inductive hypothesis that we have computed $\pi$ already for all items before $L$. This will include all elements of a strong $C$-support of $\pi(\Pi_B([L]))$, where $\Pi_B([N]) \in \tau_C^i$. If $i = 0$ we are computing $\pi$ at an atom as above, and by inductive hypothesis $\pi$ has already been computed at this atom.

If $i = 2$, extend the union of $\pi_0$ and the restriction of $\pi$ to the atoms in this strong $C$-support to a local bijection $\pi_0'$, with the restriction that no exceptions mapping from or into litters in the support are created. We can

29

enforce this restriction because we know what values of $\pi$ at local cardinals of the litters in the support are supposed to be. Apply the inductive hypothesis that the Freedom of Action theorem applies to $C \ll A$ to produce a permutation $\pi'$ extending this local bijection $\pi'_0$ without creating exceptions outside its domain. We argue that each litter $N$ in the support is mapped by $\pi'$ to the value already computed for $\pi(N)$. Suppose otherwise: let $N$ be the first counterexample in the strong support. It follows that $[N] \in \tau_D^2$ in the support is sent to the same value by $\pi$ that it is by $\pi'$ because $\pi$ and $\pi'$ agree on a $C$-support of $\Pi_D([N])$ (or in one odd case at a $C_1$-support, but this also works). If $\pi(N)$ is not the same as $\pi'(N)$ there must be exceptions of either $\pi$ or $\pi'$ at which the two maps do not agree. But in fact $\pi$ and $\pi'$ agree on all exceptions of either of the two maps (all elements of the domains of either local bijection) which lie in or are mapped into the litter $N$ ($\pi$ and $\pi'$ may disagree at some exceptions of $\pi'$ which are neither in $N$ nor mapped into $N$).

It is then clear that $\pi'(\Pi_B([N]))$ is the only possible value for $\pi(\Pi_B([N]))$

We need to verify that it doesn't matter which strong support of $x$ we use for this computation. Consider the first element of the strong support given for $x$ at which different computations of values of $\pi$ are possible. It must be a near-litter, as an atom is preceded by the litter containing it and the computation at a litter uniquely determines the value we get at each of its elements. Suppose the litter $N$ admits more than one computation. Take the support $S$ for $[N]$ extracted from the current computation and the alternative support $T$ from which the supposed alternative computation is obtained. Construct the support obtained by following $T$ with $S$ and deleting duplicates in $S$ already found in $T$. Computation along this support must give a value for $[N]$ agreeing with the computation along $T$ for values of $\pi$ on the domain of $T$ (because $T$ is considered first), and it must also give a value agreeing with the computation along $S$ for values on $S$ (because values on $S$ are unique), so the values obtained at $[N]$ must be the same (because the permutation $\pi'$ obtained as above from the long support is forced to give values at $[N]$ which agree with values computed along $S$ or $T$). Once the value at $[N]$ is determined, the value at $N$ is determined.

### 2.2.3 The power set of $\tau_A^0$ is $\tau_A^1$ in suitable FM interpretations

We show that if $A = B$ or $A \ll^* B$, then the power set of $\tau_A^0$ in the FM interpretation based on $B$-allowable permutations is $\tau_A^0$.

Clearly a set $X$ in $\tau_A^1$ has a $B$-support: $X$ is either $\bigcup Y \Delta Z$ or $(\tau_A^0 \setminus \bigcup Y) \Delta Z$, where $Y$ is a small set of litters included in $\tau_A^0$ and $Z$ is a small subset of $\tau_A^0$. Clearly $Y \bigcup Z$ is a $B$-support of $X$, and also an $A$-support.

Now suppose that a set $X \subseteq \tau_A^0$ has a $B$-support $S$: we may without loss of generality assume that $S$ is a prestrong $B$-support (prestrong because we do not have an immediate argument that it is strict).

We argue that the intersection of $X$ with any litter $L$ must be small or a co-small subset of $L$. Suppose otherwise: that $L \cap X$ and $L \setminus X$ are both large. Let $S$ be a prestrong support extending the well ordering obtained from a strong support $T$ of $X$ by appending $L$ to it if it is not already present and adding support domain elements required for $L$. Choose $a$ from $L \cap X$ and $b$ from $L \setminus X$, neither appearing in the domain of $S$. Define a local bijection swapping $a$ and $b$ and fixing each atomic element of the domain of $S$. Extend this local bijection to a $B$-allowable permutation with no exceptions not in the domain of the local bijection. This permutation will fix each litter $M$ in $S$ because it fixes a support of the local cardinal of the litter and it has no exception mapped into or out of $M$ because each of its exceptions is either fixed or mapped to another element of the same litter $L$ (in the case of $a, b$). So this allowable permutation must fix $L \setminus X$ and $L \cap X$, because it fixes a support thereof, but at the same time it clearly moves these sets. This is impossible, so $X$ must intersect any litter $L$ in a small or co-small subset of $L$.

We show that $X$ cannot cut a large collection of litters nontrivially. Suppose otherwise. Let $S$ be a prestrong $B$-support of $X$ . Let $L$ be a litter which is cut by $X$ and which does not belong to or meet the domain of $S$. Let $a$ belong to $L \cap X$ and $b$ belong to $L \setminus X$. Consider a local bijection swapping $a$ and $b$ and fixing each atomic element of the domain of $S$. Extend it to an allowable permutation with no exceptions outside the domain of the local bijection. This allowable permutation fixes each litter element of $S$ [if it failed to fix a first litter $M \in S$, it would fix $[M]$ and so would have an exception in $M$, and it has no exceptions mapped into or out of any element of $S$], and so fixes $X$. But it clearly does not fix $X$. So the collection of litters nontrivially cut by $X$ must be small.

We show that the collection of litters meeting $X$ and the collection of

litters disjoint from $X$ cannot both be large. Suppose otherwise. Let $S$ be a strong $B$-support of $X$. Choose a litter $L$ included in $X$ and a litter $M$ disjoint from $X$ and included in $\tau_A^0$ and $a \in L$ and $b \in M$, none of these belonging to the domain of $S$. Define a local bijection swapping $a$ and $b$ and fixing each atomic element of $S$. Extend it to a $B$-allowable permutation with no exceptions other than elements of the domain. This will fix each litter in $S$ (it has no exceptions which are moved and belong to elements of $S$) and so must fix $X$, but clearly does not.

From these results it follows that $X$ must have small symmetric difference from a small or co-small union of litters included in $\tau_A^0$, that is, it must belong to $\tau_A^1$.

Notice that this means that local cardinals of litters actually are subsets of the Scott cardinals of those litters.

### 2.2.4 The power set of $\tau_A^1$ is the same for $A$- and $\emptyset$-allowable permutations.

We claim that if $B = A_1$ or $A_1 \ll^* B$, the power set of $\tau_A^1$ in the interpretation based on $B$-allowable permutations is $\tau_A^2$.

It is sufficient to argue that any subset $X$ of $\tau_A^1$ with a prestrong $B$-support $S$ (prestrong because we cannot suppose it strict) also has a strong $A$-support.

And in fact this support $S'$ is easy to describe: it is simply the set of all elements of $S$ which are in a set $\tau_C^i$ with $C_i = A$ for some $i$ (i.e., $C = A$ or $C \ll A$) or $C = A_1$.

Let $\pi$ be an $A_1$-allowable permutation which fixes each element of $S'$. Our aim is to show that $\pi(X) = X$.

Let $Y$ be an element of $X$. Let $T$ be an $A$-strong support of $Y$ extending $S'$, not containing any element of $\tau_{A_1}^1$ (it is straightforward to establish that an element of $\tau_A^1$ has such a support). Define a local bijection which sends each atomic element of $T$ and each exception of $\pi$ lying in or mapped into a litter in $T$ to its image under $\pi$ and fixes each atomic element of $S \setminus S'$. We claim that the $B$-allowable permutation $\pi'$ extending this local bijection with no exceptions outside the domain of the local bijection agrees with $\pi$ on each element of $T$ and fixes each element of $S$. Note that $\pi' \circ \pi^{-1}$ fixes each atomic element of $T$ and each exception of $\pi$ lying in or mapped into a litter in $T$, which forces it to fix the local cardinal of each litter in $T$ (consider the first counterexample and the support of its local cardinal), and also each litter by restrictions on exceptions.

We verify a claim made in the previous paragraph. An element of $\tau_A^2$ can have an element of $\tau_{A_1}^1$ in its support. Consider a litter $L$ in $\tau_{A_1}^1$ and consider the union of the set of all local cardinals $[M]$ with $P_A([M]) \in L$. This is clearly a set in $\tau_A^2$ which essentially has $L$ in its support. But a set in $\tau_A^1$ cannot need such a set in its support: it has a support consisting of a small collection of elements of $\tau_{A_1}^0$, which don't generate any commitment to fixing any litter in $\tau_{A_1}^1$ and supports for elements of $\tau_{A \cup \{\delta\}}^2$ which will have litters in $\tau_A^2$ as their most complex components, which may further generate obligations concerning elements of $\tau_{A_1}^0$, their parents. This allows us to avoid the conflict between $T$ and $S \setminus S'$ which could occur if we had litters in $T$ belonging to $\tau_{A_1}^1$ which might have parents in $\tau_{A_2}^0$ whose values under $\pi$ might conflict with the need to fix litters in $S \cap \tau_{A_2}^1$.

Thus $\pi(Y) = \pi'(Y)$. Further, $\pi'$ fixes each element of $S$: all we need to

show is that it fixes litters in $S \setminus S'$. It fixes their local cardinals: consider the first counterexample and consider the action of $\pi'$ on its support; and exception discipline prevents it from moving the litters themselves because $\pi'$ has no exceptions in relevant $\tau_C^0$'s but fixed points. Thus $\pi'(X) = X$, from which it follows that $\pi(Y) \in X$ so $\pi(X) \subseteq X$. Applying the same argument to $\pi^{-1}$ shows that $\pi(X) = X$ as desired.

### 2.2.5 The size of sets $\tau_A^2$ is $\mu$

The map $\Pi_A$ cannot be defined unless $\Pi_{A \setminus A_2}$ can be defined, which requires that $\tau_{(A \setminus A_2)_1}^0$ be of size $\mu$ in the ambient set theory (true) and that $\tau_{(A \setminus A_2) \cup \{\delta\}}^2$ be of size $\mu$ in the ambient set theory, where $\delta < \mathtt{min}(A)$: for this it is sufficient that $\tau_{\mathtt{min}(A),\delta}^2$ be of size $\mu$ for each $\delta < \mathtt{min}(A)$, since this set is the same size (a fact witnessed by a $\chi$ map). This gives us enough information to establish that $\tau_A^2$ *exists*. To complete an argument by induction that everything works correctly, we need to show further that $\tau_A^2$ is of size $\mu$ in the ambient set theory.

There are $\mu$ subsets of size $< \kappa$ of a set of size $\mu$ (the cofinality of the strong limit cardinal $\mu$ being at least $\kappa$). There are $\mu$ litters in any $\tau_B^1$ (obvious). There are $\mu$ small sets of these litters and there are $\mu$ small subsets of $\tau_B^0$ as already noted, so there are $\mu$ elements of $\tau_B^1$, by the description of elements of $\tau_B^1$ already given. There are $< \mu$ finite subsets of $\lambda$. So it follows that there are $\mu$ supports.

We introduce another special kind of support. A nice $A$-support of $X$ is an $A_1$-support $S$ of $X$ such that

1. if its domain meets a $\tau_C^i$, then $C \ll A, C = A$ or $C = A_1$,

2. and distinct elements of sets $\tau_C^1$ in the domain are disjoint,

3. and if an element $x$ of the domain belongs to another element $N$ of the domain, $N$ precedes $x$ in the support,

4. and if a near-litter $N \in \tau_C^1$ with $C \neq A_1$ appears in the support, a strict support (which will be nice) of $\Pi_C([N^\circ])$ is included in the segment determined by $N$.

Note that a strong $A$-support is a nice $A$-support, and the image under an $A_1$-allowable permutation of a nice $A$-support is a nice $A$-support.

If $S$ is a nice $A$-support of $x$, we define the coding function $\xi_{x,S}$ so that if $\pi$ is an $A_1$-allowable permutation, $\xi_{x,S}(\pi(S)) = \pi(x)$. Note that images of nice $A$-supports under $A_1$-allowable permutations are nice $A$-supports (this is not true of our other special flavors of support) and that if $\pi(S) = \pi'(S)$ then $\pi^{-1} \circ \pi'(S) = S$, so $\pi^{-1} \circ \pi'(x) = x$, so $\pi(x) = \pi'(x)$ ($\pi$ and $\pi'$ $A_1$-allowable). The coding function is a function from the orbit of $S$ under allowable permutations to the orbit of $x$ under allowable permutations.

With a nice $A$-support $S$ we associate a combinatorial object $\texttt{format}(S)$, a well-ordering of the same type as $S$, constructed by replacing items in $S$ as follows:

1. Replace an atoim in $\tau_C^0$ in position $\alpha$ in $S$ with $(\alpha, 1, C, \beta)$ where either the object at position $\beta$ in $S$ is a near-litter containing the atom or there is no near-litter in the domain of $S$ containing $x$ and $\beta = \kappa$.

2. Replace a near-litter $N$ in position $\alpha$ with set parent $x$ in $\tau_C^2$, $C \neq A_1$, with $(\alpha, 2, C_1, \xi_{x,S'})$ where $S'$ is the maximal nice $C$-support of $x$ included in the segment determined by $N$.

3. Replace a near-litter $N$ in position $\alpha$ belonging to $\tau_C^1$, $C \neq A_1$ and parent an atom with $(\alpha, 3, C, \beta)$, where the parent is at position $\beta$ in $S$.

4. Replace a near-litter $N$ in position $\alpha$ belonging to $\tau_{A_1}^1$ with $(\alpha, 4)$.

Two nice $A$-supports have the same format if and only if there is an $A_1$-allowable permutation which takes one to the other. The reason for this is that given $S$ and $T$ with $\texttt{format}(S) = \texttt{format}(T)$ we can construct a local bijection whose extension to an $A_1$-allowable permutation will take $S$ to $T$. The local bijection takes atoms in each position in $S$ to atoms in the same position in $T$. If $N$ is a near-litter in $S$ and $N_*$ is the near-litter in $T$ in the same position, choices made for values of the local bijection earlier (if made successfully) already ensure that $[N]$ will be mapped to $[N_*]$; we need to extend the local bijection to ensure that $N$ is mapped exactly to $N_*$. Elements of $N \setminus N^\circ$ need to be mapped by the local bijection to elements of $N^*$ and elements of $N_* \setminus N_*^\circ$ need to be mapped by the inverse of the local bijection to elements of $N^\circ$, and elements of $N^\circ \setminus N$ need to be mapped by the local bijection to non-elements of $N^*$ and elements of $N_*^\circ \setminus N_*$ need to be mapped by the inverse of the local bijection to non-elements of $N^\circ$; the local bijection then needs to be filled out to complete orbits for each item without introducing further exceptions which belong to any of the litters $N^\circ$ for $N \in S$ or $N_*^\circ$ for $N_* \in T$ [this condition can be met because the intended targets of $[N]$ under the permutation to be constructed and $[N_*]$ under the permutation to be constructed can be read from $S$ and $T$; so one can be sure that things in the appropriate litters are mapped to elements of appropriate litters by the local bijection and its inverse while filling out the orbits.

Thus the formats exactly code the orbits of nice $A$-supports under $A_1$-permutations.

Our aim is to show that there is a set of coding functions of size $< \mu$ the union of whose ranges includes all elements of $\tau_B^2$'s. Since there are only $\mu$ supports, this establishes that the size of each $\tau_B^2$ is no more than $\mu$. Further, since $\tau_B^2$ certainly contains $\mu$ double singletons of atoms in $\tau_B^0$, the size of $\tau_B^2$ is exactly $\mu$.

Fix a nonempty subset $A$ of $\mu$. Assume that there is a set of size $< \mu$ of coding functions whose ranges cover all $\tau_B^2$ for $B \ll^* A$. We indicate how to give descriptions of $< \mu$ coding functions whose ranges will cover $\tau_A^2$.

First, we give descriptions of $< \mu$ coding functions which cover $\tau_A^1$. Let $X$ be a subset of $\tau_A^1$. We determine a coding function for $X$ by taking a small collection of litters whose union or the complement of whose union has small symmetric difference from $X$ and the atoms in that small symmetric difference, and choosing nice supports for each of these items. Each of these items $x$ is either in $\tau_B^2$ for some $B \ll A$ and has parent a value of a coding function which may be written $\xi_{x,S_x}$ in the collection of $< \mu$ coding functions given by ind hyp ($S_x$ is actually a coding function based on a support with the additional litter but based on a coding function for the parent in the covering collection of coding functions) or is in $\tau_{A_1}^0$ and has nice support $S_x$ whose domain is just that atom and the litter. Merge all the supports $S_x$ into a support $S_X$ and $\xi_{X,S_X}$ is a coding function for $X$. There are $< \mu$ possible formats for $S_X$, with two sets $X$ with the associated support $S_X$, and so $< \mu$ coding functions generated in this way: the reason is that the format of $S_X$ is a merger of a small list of formats taken from a set of $< \mu$ possible formats, and $< \mu$ possible formats can be obtained in this way (the cofinality of $\mu$ being at least $\kappa$). Merging formats is tricky to describe formally because all internal ordinals pointing to positions in the merged formats need to be changed to address the new positions of the items pointed to in the merged supports, but this doesnlt affect the combinatorics of saying how many merged formats there are.

Now we describe coding functions for general elements of $\tau_A^2$. Let $X$ be an element of $\tau_A^2$ with nice support $S$. For each element $Y$ of $S$, choose a coding function $\xi_{Y,S_Y}$ from the set of $< \mu$ coding functions described above and generate a support $S_Y^*$ from the merger of $S$ followed by $S_Y$. The desired coding function is determined by the format of $S$ and the set of supports $S_Y^*$ generated as above. The idea is that any support $T$ with the same format as $S$ is mapped to the collection of all $\xi_{Y,S_Y^*}(T^*)$ where $T^*$ extends $T$ and is

37

in the domain of the coding function. We need to show that this operation sends $S$ to $X$. Each element of $Y \in X$ is $\xi_{Y,S_Y^*}(S_Y^*)$ where $S_Y$ is a suitable $S^*$. Now we show that $\xi_{Y,S_Y}(S^*)$ where $S^*$ extends $X$ actually belongs to $X$: the trick is to observe that we can define a local bijection from $S_Y^*$ to $S^*$ which fixes all items in $S$ and the allowable permution extending which will map $S_Y^*$ to $S^*$ and so $\xi_{Y,S_Y^*}(S_Y^*)$ to $\xi_{Y,S_Y}(S^*)$, and this map fixes $X$ because it fixes every element of $S$.

We complete the counting argument. There are two components, the format of the support $S$ and the set of functions $\xi_{Y,S_Y^*}$. There are $< \mu$ possible functions $\xi_{Y,S_Y^*}$ for each $S$, determined by the collection of $< /mu$ coding functions $\xi_{Y,S_Y}$ and the small collection (for each coding function) of possible schemes of relationship of elements of $S$ with elements of a support with the format of $S_Y$ in a merged support: the second component is a set of coding functions from a set of $< \mu$ possible coding functions, so there are $< \mu$ possible second components because $\mu$ is strong limit.

We need to argue that there are $< \mu$ possible formats for the support $S$. A support is built out of a small number of components, ordinals and finite subsets of ordinals which are taken from sets of small cardinality, $\xi_{x,S}$'s for $x \in \tau_B^2$ for $B \ll^* A$, which are taken from the covering collection of coding functions which is of size $< \mu$, and a small collection of $\xi_{x,S}$'s for $x \in \tau_A^2$. These last might seem problematic, but the solution is to build the formats in stages indexed by $\kappa$: at each stage one builds formats using only $\xi_{x,S}$'s for $x \in \tau_A^2$ built at lower stages. Clearly at each stage there are $< \mu$ formats introduced, and there are $< \mu$ in all because $\kappa < \mu$. It should also be clear just from the way formats are constructed that a $\xi_{x,S}$'s for $x \in \tau_A^2$ appearing in the format of $T$ will be constructed at an earlier stage than the format of $T$.

Finally, one obtains a covering set of coding functions for all $\tau_A^2$'s because one proceeds through $|\lambda| < \kappa$ steps indexed by sets $A$.

# 3   Conclusions