An outline of a proof of the consistency of New Foundations

M. Randall Holmes
Boise State University

Given 6/18/2021 at BEST (up to p. 25),
updated 6/23
the full set of slides really does contain an
outline of the proof, which I will be refining;
improved versions of these slides will be
accessible on my web page at

`https://randall-holmes.github.io/Nfproof/nfoutlineslides.pdf` .

There could easily be slips in this
presentation; I'm trying out a somewhat
different structure for the argument.

The Problem

In 1937, W. v. O. Quine, a notable Americal philosospher and logician, proposed what is perhaps the most streamlined possible version of Russell's theory of types from Principia Mathematica, in a paper titled New Foundations for Mathematical Logic, from which the theory is usually called NF (for "New Foundations").

New Foundations is a variation on TST, the simple typed theory of sets. This is the sorted theory with equality and membership as primitive predicates, the sorts being indexed by natural numbers and the legal forms for atomic sentences neatly summarized by the schemata $x^i = y^i$, $x^i \in y^{i+1}$, and axioms of extensionality and comprehension exactly as in naive set theory (comprehension being restricted by the rules for formation of sentences: $\{x : x \notin x\}$ is not provided by an instance of comprehension because a sentence of the shape $x \in x$ is not well-formed, no matter what the type of x is).

It is an interesting historical note that TST is **not** the type theory of Principia Mathematica, though a summary of TST is often presented as an account of PM by careless writers. TST appears to be described first by Tarski around 1930. The obstruction to Russell and Whitehead presenting their theory in this way is that they had no idea how to implement the ordered pair using sets, so the type system of PM is a complicated system of relation types, further complicated by predicativity considerations. Norbert Wiener appeared to have TST in mind when he presented the first set theoretical definition of the pair in 1914, but he did not give a formal description.

TST isn't the simple type theory of Ramsey, either: that is the theory of n-ary relations of PM without the predicativity restrictions, and so without need of the axiom of reducibility which subverts the predicativity restrictions.

NF is motivated by observing the phenomenon of systematic ambiguity in TST, which Russell had already noted in the system of PM. This symmetry is much more striking in TST. Provide an injective operation $x \mapsto x^+$ on variables which raises type by one. Let ϕ^+ be the result of replacing each variable x in ϕ with x^+ throughout. Then ϕ^+ is a theorem of TST if ϕ is a theorem (the converse is not true: more can be proven about higher types), and any object defined in set builder notation in a form $\{x : \phi\}$ has an exact analogue $\{x^+ : \phi^+\}$ in the next higher type (and this can be iterated).

Quine's proposal was that it seems reasonable with such a high degree of symmetry to suppose that the types are simply **the same**: the resulting theory is a single sorted theory with equality and membership, with axioms of extensionality and a comprehension scheme consisting of those assertions "$\{x : \phi\}$ exists" which can be obtained from instances of the comprehension scheme of TST by ignoring distinctions of type between the variables.

It is traditional to give an account of the comprehension scheme of NF which does not mention the rules of sentence formation of another theory. A function σ from variables to natural numbers is called a stratification of a formula ϕ if each subformula '$x = y$' of ϕ satisfies $\sigma(\text{'}x\text{'}) = \sigma(\text{'}y\text{'})$ and each subformula '$x \in y$' of ϕ satisfies $\sigma(\text{'}x\text{'}) + 1 = \sigma(\text{'}y\text{'})$. A formula ϕ is said to be stratified iff there is a stratification of ϕ. The comprehension axiom of NF can be presented in the form "$\{x : \phi\}$ exists if ϕ is stratified". This has been criticized as a syntactical trick. It is worth noting that stratified comprehension is equivalent to a finite subset of its instances, so in fact it can be expressed in a way which makes no reference to types at all. A finite axiomatization can be developed by analogy with the finite axiomatization of the class comprehension axiom scheme of von Neumann-Gödel-Bernays class theory.

We will not allow ourselves to be distracted too much about the oddities of the way the world looks in this theory. The universal set exists and sets make up a Boolean algebra under the usual operations. The Frege natural numbers exist and are the natural implementation of $\mathbb{N}$. More generally, Russell-Whitehead cardinals and ordinals exist and are the natural implementations of cardinal and ordinal numbers. The Russell paradox is trivially avoided ($x \notin x$ is not a stratified formula). The Cantor paradox is avoided because the form of Cantor's theorem in TST is $|\iota``A| < |\mathcal{P}(A)|$ $[\iota = (x \mapsto \{x\})]$ instead of the ill-typed $|A| < |\mathcal{P}(A)|$. $|A| < |\mathcal{P}(A)|$ is a well-formed assertion in the language of NF, but clearly false if $A = V$, the universal set. $|\iota``V| < |\mathcal{P}(V)|$ is provable, so the singleton operation is not implemented by a set function: the collection of singletons is smaller than the universe in spite of the obvious external bijection between these sets. The way in which NF avoids the Burali-Forti paradox is fascinating but would take us too far afield.

The first indication that NF is not a harmless notational variation of TST was Specker's 1954 result that the Axiom of Choice is refutable in NF. We won't discuss the details of the proof of this result at all, but the bare fact may serve to provide a hint of the motivation behind the approach we took to the problem.

At this point if not before it was clear that there is a **problem** of consistency of NF relative to a set theory in which we have confidence. That NF proves the negation of Choice shows that NF is at least as strong as TST $+$ Infinity: there is no evidence that NF is any stronger than TST $+$ Infinity, which has the same strength as Zermelo with Separation restricted to bounded formulas.

In 1962, Specker proved that NF is equiconsistent with TST with the Ambiguity Scheme asserting $\phi \leftrightarrow \phi^+$ for each closed formula ϕ. This result is not entirely trivial to prove, though it is strongly suggested by Quine's original motivation in defining this theory.

In 1969, R. B. Jensen proved that NFU (New Foundations with urelements), the theory obtained by modifying NF by weakening extensionality to apply only to objects with elements, is consistent and moreover is consistent with Infinity and with Choice (it is consistent with stronger axioms of infinity as well, and this is clear from Jensen's original paper).

Specker's result can be generalized to show that NFU is equiconsistent with TSTU + Ambguity, where TSTU has extensionality weakened to allow many objects with no elements in each positive type.

Jensen's consistency proof shows that the ways that NF avoids the Russell, Cantor, and Burali-Forti paradoxes work in a set theory we can have confidence in. If NF were inconsistent, it would fall prey to a different paradox. Specker's proof can be understood as showing that NF + Choice falls victim to a different paradox of set theory.

It is another interesting matter of history to note that Quine discusses his choice of strong extensionality as an axiom for NF: he claims it can be harmlessly motivated by the identification of urelements u with their singletons $\{u\}$, which is an actual mathematical error in a context with stratified comprehension. It would take us too far afield to discuss the details, but it is worth noting for the benefit of anyone who might remember this from looking at the original paper.

The second step to the solution: tangled type theory and tangled webs (Holmes 1995)

The first step to the solution is Jensen's 1969 proof of the consistency of NFU. I'll be able to retrospectively give an account of a version of Jensen's proof after I give an account of my 1995 results.

Tangled type theory

TTT (tangled type theory) is a theory with sorts indexed by ordinals less than a limit ordinal λ (which may be taken as ω but does not have to be) with formation rules for atomic sentences $x^\alpha = y^\alpha$ and $x^\alpha \in y^\beta$ for $\alpha < \beta$.

For any strictly increasing sequence s of ordinals less than λ and formula ϕ in the language of TST, define ϕ^s as the formula in the language of TTT obtained from ϕ by replacing each variable x^i in ϕ with x^{s_i}. (We can make exact sense of this by requiring in both theories that each variable is of the form $\mathbf{x}_n^\tau$ where n is a natural number index and τ is a type).

The axioms of TTT are the formulas ϕ^s where ϕ is an axiom of TST and s is any strictly increasing sequence of ordinals less than λ.

For mental hygiene, I strongly suggest not thinking about what the world looks like in TTT. Focus simply on the formal proof which follows that TTT is exactly as strong as NF.

it is important to note that TTT is **not** cumulative type theory. Each type is being presented as a "power set" of *each* lower type, which is very strange; the types can be thought of as disjoint as is usual in TST, though the language doesnt permit us to say this.

From a model of NF one immediately obtains a model of TTT by using a copy of the model of NF to implement each type in the model of TTT, the membership relations between copies being determined by the membership relation of the model itself in the obvious way.

Suppose that we are given a model of TTT.

Let Σ be a finite set of sentences in the language of TST. Let n be a strict upper bound on types mentioned in formulas in Σ. We use Σ to define a partition of the collection $[\lambda]^n$ of n element subsets of λ: the compartment in which $A \in [\lambda]^n$ is put is determined by the truth values of formulas ϕ^s for $\phi \in \Sigma$ and $\mathrm{rng}(s\lceil n) = A$. This partition has an infinite homogeneous set H which includes the range of a strictly increasing sequence h. The model of TST determined in the obvious way by the sequence h of types taken from λ (ϕ holds in this model iff ϕ^h holds in our model of TTT) satisfies TST + (Ambiguity restricted to formulas in Σ). But this implies directly that TST + Ambiguity is consistent by compactness, so NF is consistent by the results of Specker.

This argument is motivated by Jensen's original proof of Con(NFU) and can be reverse engineered to a proof of Con(NFU). The idea is that TTTU has an obvious model: let type α $(\alpha < \lambda)$ be implemented by V_α, the level of the cumulative hierarchy indexed by α, with $x^\alpha \in y^\beta$ taken as meaning $x^\alpha \in y^\beta \wedge y^\beta \in V_{\alpha+1}$, where x^α ranges over V_α and y^β ranges over V_β in the familiar set theoretical universe. Notice that all elements of $V_\beta \setminus V_{\alpha+1}$ are being treated as urelements here. This can further be seen to show that NFU has models exhibiting stratified versions of whatever mathematical statements we might expect to be true in a suitably chosen level of the cumulative hierarchy: NFU is only superficially different from ordinary set theory, in a sense which we have discussed at length elsewhere.

TTTU has natural models which we have just described. Natural models of TTT are quite another matter. The difficulty is that each type is being interpreted as a "power set" of *each* lower type, not just an immediate predecessor, and simple considerations of cardinality make it clear that something very strange has to be going on for this to be achieved.

By contrast, in the argument for Con(NFU) each type is being interpreted as the power set of each lower type...plus lots of urelements in each case. And we have indicated how to do this straightforwardly.

We modify (as it were "unfold") this picture to something which may appear achievable if still weird in the ordinary set theoretical universe.

We briefly discuss what we mean by "unfolding" a model of TTT. We hope that on this brief explanation, the audience will recognize a tangled web as something related to unfolding a model of TTT.

The problem in TTT is that for types $\alpha < \beta < \gamma$ we see type β as consisting of subsets of type α (all the ones we can define with appropriate type discipline) and type γ consisting of subsets of type β similarly. This is unexceptionable. But we also see type γ as consisting of subsets of type α (all the ones we can define with suitable type discipline!); so it cannot be the case that each of these understandings of a type as the power set of a lower type is honest, simply by Cantor's theorem.

The idea of the unfolding is that we replace each type α with a collection of types labelled by subsets A of λ with α as minimal element.

The power set of the type labelled by the set A (secretly a copy of type $\alpha = \min(A)$) is then taken to be the type labelled by $A \setminus \{\min(A)\}$ − so each type has a designated power type. We use the notation A_1 for $A \setminus \{\min(A)\}$ and A_{n+1} for $(A_n)_1$. We could use countable sets (there is no reason to use a subset of λ with order type $> \omega$), but we in fact use finite sets (because no statement in the language of type theory has to refer to infinitely many types).

If we take any type indexed by a set A and consider the model of an initial segment of type theory formed by types $A, A_1, A_2, A_3, \ldots, A_{n-1}$, and consider the idea that type A is secretly understood to be a copy of type $\min(A)$ in TTT, we should get the idea that the theory of this initial segment should depend only on the smallest n elements of the set A (in our eventual construction we let it depend on the first $n + 1$ elements for technical reasons).

The formal definition following embodies these ideas.

Oh what a tangled web we weave. . .

We work in ordinary set theory without choice and with extensionality weakened to allow a set of atoms, in which we can use the Scott definition of cardinal (this requires nothing more than bounded Zermelo set theory plus the condition that every set belongs to a rank of the cumulative hierarchy (level 0 of the hierarchy being the set of atoms)).

Let λ be a limit ordinal. For any nonempty finite subset A of λ, define A_1 as $A \setminus \{\mathtt{min}(A)\}$. Define A_0 as A and A_{n+1} as $(A_n)_1$.

TST$_n$ is the subtheory of TST in which only types less than n are used (so there are n types, the lowest one being 0). A natural model of TST$_n$ is one in which the lowest type is a set X, each type $i < n$ is implemented by $\mathcal{P}^i(X)$, equality within each type is the equality of the ambient set theory, and membership of objects of each type in objects of the next type is the membership relation of the ambient set theory. Esthetically one might prefer that the types be disjoint and this can be arranged but is not actually needed.

It is an important fact that the theory of a natural model of TST$_n$ depends only on n and the cardinality of the set implementing type 0.

A *tangled web* is a function τ from nonempty finite subsets of λ to cardinals with the following properties:

naturality: for any A with $|A| \geq 3$, $2^{\tau(A)} = \tau(A_1)$. $|A| \geq 2$ would seem more natural but our construction motivates the form given here.

elementarity: for any A with $|A| \geq n + 1$, the theory of a natural model of TST_n with type 0 implemented by a set of size $\tau(A)$ depends only on $A \setminus A_{n+1}$ (the set consisting of the $n + 1$ smallest elements of A). Using n rather than $n + 1$ elements would seem more natural but our construction motivates the form given here.

The existence of a tangled web implies the consistency of NF. The proof is very similar to the proof given above for tangled type theory.

Let Σ be a finite set of sentences in the language of TST. Let n be a strict upper bound on types mentioned in formulas in Σ. We use Σ to define a partition of the collection $[\lambda]^{n+1}$ of $n + 1$ element subsets of λ: the compartment in which $A \in [\lambda]^{n+1}$ is put is determined by the truth values of formulas $\phi \in \Sigma$ in natural models of TST_n with base types of size $\tau(B)$ where $B \setminus B_{n+1} = A$. This partition has a homogeneous set H of size $n + 2$.

We observe that a natural model of TST_{n+1} with base type of size $\tau(H)$ has the same truth values for Σ as the model of TST_{n+1} whose base type is the set implementing type 1 in the previous model whose size is $2^{\tau(H)} = \tau(H_1)$, by homogeneity of H with respect to the indicated partition. The punchline is that that TST_{n+1} is consistent with ambiguity for sentences in Σ, so TST is consistent with ambiguity for sentences in Σ, so $\mathsf{TST} + \mathrm{Ambiguity}$ is consistent by compactness.

It is worth noting somewhere, and it might as well be here, than both TTT and ordinary set theory minus Choice with a tangled web disprove Choice, in a manner which can be developed by analogy with Specker's disproof of Choice in NF.

(added after the talk)

Perhaps ill-advisedly for mental hygiene, we discuss the extent to which TTT can see tangled webs.

The natural model of TST_n with a set X as base type is actually a notion expressible (deviously) in TST: the sequence of sets $\mathcal{P}^i(X)$ is not definable in a stratified way, but the finite sequence of sets $\iota^{n-1-i}\text{``}\mathcal{P}^i(X)$, where i ranges from 0 to $n-1$, is a set and its membership relations (different relations between each successive pair of types) are well-typed (relations on suitably iterated singletons induced by repeated applications of the singleton operations).

Thus in TTT we can express any finite fragment of the system of types as a system of sets. In a model of TTT in which natural

numbers are standard, the properties of a tangled web will hold for the cardinals of the types in such a system. In a model of TTT in which any subcollection no larger than λ of a type is a set, a full tangled web for any $\lambda_0 < \lambda$ (where λ makes up the types of the model of TTT) is present.

Unfolding happens automatically. Consider three types α, β, γ. Two different representations of type γ appear via the set of singletons (type α) of singletons (type β) of type γ objects and the set of singletons (type α) of type γ objects. These sets are not the same size in the internal sense of the model of TTT. The power set of the second set is the same size as type α, and the double power set of the second set is the same size as type α (as seen in any sequence of types above type α), so they must be of distinct sizes. This is general: types are being coded by cardinals, and a given type will have many distinct codes at most higher types.

The final move: constructing a tangled web

There is a version of the proof which involves constructing a model of tangled type theory directly. Tangled type theory is bewildering; the earliest versions of the argument took the approach of constructing a tangled web, and that is what we will do here.

The argument is via construction of a Frankel-Mostowski permutation model in ZFA. Recall that the Fraenkel-Mostowski model technique was originally developed to argue that ZFA was independent of Choice; the failure of Choice in NF suggests that such a technique might be used here.

There are some cardinal invariants of the construction.

λ is a limit ordinal.

For finite subsets A of λ with at least n elements, we define A_n as above.

κ is a regular uncountable cardinal. A set of size $< \kappa$ is called small; all other sets are called large.

μ is a strong limit cardinal of cofinality greater than λ or κ ($\geq$ may suffice).

We work in ZFA with μ atoms.

We specify a well-ordering $\ll$ on finite subsets of λ. It is uniquely specified by three conditions:

If $A \neq \emptyset$, then $A \ll \emptyset$

If $\max(A) < \max(B)$, then $A \ll B$

If $\max(A) = \max(B)$, then $A \ll B$ iff $A \backslash \{\max(A)\} \ll B \setminus \{\max(B)\}$.

Note that $A \ll A_i$ if $i > 0$: downward extensions of a set appear before the set in this order.

There are μ atoms, partititioned into sets τ_A^0 for each nonempty finite subset A of λ; each of these sets is of size μ. These sets are called *clans*. The notation for τ_A^0 in older versions of this argument is `clan[A]`.

Each set τ_A^0 is partitioned into sets of size κ called *litters*. A subset of a τ_A^0 with small symmetric difference from a litter is called a *near-litter*. For any litter L, the set of all near-litters with small symmetric difference from L is called the local cardinal of L, written $[L]$.

We define τ_A^1 as the collection of all subsets X of τ_A^0 for which there is a small set Y of litters included in τ_A^0 such that either $X \triangle \bigcup Y$ is small or $X \triangle (\tau_A^0 \setminus \bigcup Y)$ is small: that is, X has small symmetric difference from a small or co-small union of litters included in τ_A^0.

The value $\tau(A)$ of our tangled web will be the size in a suitable Fraenkel-Mostowski model of a collection τ_A^2 of subsets of τ_A^1 which we will describe.

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

We extend the action of χ_α 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 χ_A is to provide a way to identify elements of sets τ_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", in which all types indexed by sets with the same minimal element are to be understood as secretly the same.

We will construct for each A a set $\tau_A^2 \subseteq \mathcal{P}(\tau_A^1)$. All these sets are of cardinality μ. [In the eventual FM model, the cardinality of τ_A^2 will be $\tau(A)$, where τ is the desired tangled web.]

We define K_A as the collection of local cardinals of litters included in τ_A^0. 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 Π_A as $\chi_{\max(A)}^{-1}(\Pi_{A\setminus\{\max(A)\}})$. It follows that Π_A is a bijection from K_A to the union of $\tau_{A_1}^0$ and the union of all τ_B^2 for which $B_1 = A$.

Traditional terminology which I might slip into if I am not careful is to, when x is an atom in the clan τ_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 $N \in [L]$.

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

An A-allowable permutation is a permutation of atoms whose action fixes each τ_B^0, fixes each K_B (so it always maps litters to near-litters in the same clan), and fixes Π_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 π such that $\pi(S) = S$ also satisfies $\pi(X) = X$. A strict A-support is one for which if its domain meets τ_B^i, there is i such that $B_i = A$ (so certainly $B \ll A$ or $B = A$).

An object X has prestrong A-support S iff X has A-support S and S is a strict A-support and each atom in the domain of S which is not in τ_A^0 belongs to a near-litter in the domain of S preceding it in S and for each near litter $N \in \text{dom}(S)$ belonging to τ_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 $i = 0, 2$, unless $C = A$, and each near-litter in the domain of S is a litter.

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

The collection τ_A^2 consists exactly of those subsets of τ_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 τ_A^2, mod the choice of the maps K_B for $B \ll A$ and $B = A$, and an annoying refinement described in an imminent slide, as long as we can verify that τ_A^2 is of size μ in the ambient set theory.

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

The basic Fraenkel-Mostowski construction

Each of the sequences of sets τ_A^0, τ_A^1, τ_A^2 is an example of the same basic FM construction which I want to describe intuitively at this point to give the audience an idea of what is going on.

The family of permutations used (considering just this small part of the model) are those which map litters to near-litters (so all but a small number of atoms in any given litter L are mapped to the same litter $\pi(L)^\circ$ — but there may be a small collection of exceptions) and which further fix some map Π from the local cardinals to some unspecified remote part of the universe. Sets in the FM permutation are those which have small supports made up of atoms and near-litters relative to this class of permutations.

In this model, the litters will be κ-amorphous (all their subsets are small or co-small) and it takes only a moderate amount of work (appearing later in these slides but probably not covered in today's talk) to show that the subsets of τ_A^0 in this model are exactly the sets with small symmetric difference from small or co-small collections of litters, i.e., exactly the elements of τ_A^1.

Local cardinals $[L]$ will be sets in τ_A^2 in the model. These sets are all as it were indistinguishable from an external standpoint. It is important to notice that while each litter is a set in the FM model, the collection of litters is not a set! The effect of the invariance of the map Π under the permutations used is to allow us to control what sets of local cardinals there are, and equivalently, what set unions of sets of local cardinals belong to τ_A^2. The underlying idea here is that there is a terrific amount of freedom in deciding what sets of these indistinguishable local cardinals to have in the model.

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 μ. 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, $[M] <^2_{\alpha,\beta} [L]$).

A consequence of this is that every element of any τ_A^2 has an A-strong support with the further property that for for each $L \in \tau_A^1$ 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 τ_A^1, we have $[M] <_A^2 [L]$.

A further consequence is that any strict A-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 ω 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 τ_B^1 will either be in τ_B^1 and earlier in the well-ordering $<_B^2$, or will be in a τ_C^2 with $C \ll B$. [I'm well aware that

demonstrating that this works requires care]. The argument justifies the stronger expectation of strong supports that a strong A-support contains for each litter L in τ_B^1 with $\Pi_B([L])$ in τ_C^2, a strong C-support (and so a prestrong B-support) of $\Pi_B([L])$.

The collection of all objects with $\emptyset$-supports (hereinafter supports) is a model of ZFA by the usual results about FM constructions; the collection of all objects with A-supports is similarly a model of ZFA. It is useful and important to note that any small subset of the model is included in the model, because a union of a small collection of supports is a support.

τ_A^1 is the power set of τ_A^0 in the FM model defined by B-permutations for $B = A$ or for any B with $A \ll B$ (this is a statement which requires verification, but should not be difficult to believe). Thus the old notation for this set was $\mathcal{P}_*(\mathtt{clan}[A])$.

τ_A^2 is the power set of τ_A^1 in the FM model determined by A_1-allowable permutations (this is strongly suggested by the way it is defined) and so is the double power set of τ_A^0. We make the claim to be verified that the subsets of τ_A^1

with supports are the same as the subsets with A_1-supports, and in fact the same as those with strong A-supports, so in fact τ_A^2 is the power set of τ_A^1 in the FM model determined by all allowable permutations. Thus the old notation for this set was $\mathcal{P}_*^2(\mathtt{clan}[A])$.

We verify that (subject to claims which need to be verified later) we can show that $\tau(A) = |\tau_A^2|$ defines a tangled web in the FM model determined by all allowable permutations. Please note that cardinalities and exponentiation maps on cardinals in the argument for the elementarity property which follows are those of the FM model.

Obviously $|K_A| \le |\tau_A^2|$, since elements of K_A are elements of τ_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 $2^{|K_A|} \le \tau_A^2$, because subsets of K_A are in one to one correspondence with their set unions, which are elements of τ_A^2, because K_A is a pairwise disjoint collection. Because of the existence of the map Π_A, we have $|\tau_{A_1}^0| \le |K_A|$ and $|\tau_B^2| \le |K_A|$ when $B_1 = A$. We define $\exp(\kappa) = 2^\kappa$.

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 τ.

The natural model of TST_n with base type τ_A^2 is sent by the composition of χ_α's determined by the elements of A_{n+1} to the natural model of TST_n with base type $\tau_{A\backslash A_{n+1}}^2$, and the χ_α'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 $|A| \geq n+1$. 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 $\tau_{(A\backslash A_{n+1})_i}^2 = \tau_{A_i\backslash A_{n+1}}^2$, and independently of the value of i the same composition of χ_α's serves as an external isomorphism. This verifies the elementarity property of τ.

This is an outline of how the proof works. What remains is the careful analysis of the way allowable permutations work which serves to verify that each set τ_A^2 is of size μ, that the power set of τ_A^0 in the FM models is τ_A^1, and that the subsets of τ_A^1 in the model determined by A-allowable permutations are the same as those in the model determined by all allowable permutations. What is required is results showing that allowable permutations act quite freely, and that is a further story.

The rest of the story: careful analysis of allowable permutations

The rest of the argument hinges on very careful analysis of allowable permutations and supports.

Everywhere here we assume that a strong A-support includes for each $L \in \tau_B^1$ in its domain with $\sqcap_B([L]) \in \tau_C^1$ and $B \neq A_1$ a strong C-support of $\sqcap_B([L])$, a detail which is justified by detail of the argument that each A-support can be extended to a strong A-support.

For any near litter N, we define N° as the litter with small symmetric difference from N. If π is an allowable permutation, we say that an atom x is an exception of π 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 τ_A^0 to elements of the same τ_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 π_0 whose domain meets no τ_C^0 with $A \ll C$ can be extended to an A-allowable permutation π with the property that each exception of π is either fixed by π or belongs to the domain of π_0.

We commence proving the Freedom of Action theorem. Fix a local bijection π_0 and a finite subset A of λ.

Specify a well-ordering $<_L$ of type κ 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 a strong support of x.

If x is in the domain of π_0, $\pi(x) = \pi_0(x)$.

If x is in a τ_B^0 with $A \ll B$, and not in the domain of π_0, $\pi(x) = x$. Alternatively, π could be made to agree with an arbitrary π' extending π_0 and sending local cardinals to local cardinals at such values.

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

It should be evident that what we have said already enforces that π has no exceptions outside the domain of π_0.

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

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

If $i = 2$, extend the union of π_0 and the restriction of π to the atoms in this strong C-support to a local bijection π_0', with the restriction that no exceptions mapping from or into litters in the support are created. Apply the inductive hypothesis that the Freedom of Action theorem applies to $C \ll A$ to produce a permutation π' extending this local bijection π_0' without creating exceptions outside its domain. We argue that each litter N in the support is mapped

by π' 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 π that it is by π' because π and π' 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 π or π' at which the two maps do not agree. But in fact π and π' 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 (π and π' may disagree at some exceptions of π' 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 π 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 π 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 π' 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.

The power set of τ_A^0 is τ_A^1 in suitable FM interpretations

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

Clearly a set X in τ_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 τ_A^0 and Z is a small subset of τ_A^0. Clearly $Y \cup Z$ 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, and so has a strong B-support S.

We argue that the intersection of X with any llitter 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 strong support extending the well ordering obtained from a strong

support T of X by appending L to it if it is not already present. 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 strong 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, 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 τ_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 τ_A^0, that is, it must belong to τ_A^1.

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

The power set of τ_A^1 is the same for A_1- and $\emptyset$-allowable permutations.

The power set of τ_A^1 in the interpretation based on A-allowable permutations is τ_A^2. We claim that if $A \ll B$, the power set of τ_A^1 in the interpretation based on B-allowable permutations is also τ_A^2.

It is sufficient to argue that any subset X of τ_A^1 with a strong B-support S 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 τ_C^i with $C_i = A$ for some i or $C = A_1$.

Let π 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 $\tau^1_{A_1}$ (it is straightforward to establish that an element of τ^1_A has such a support). Define a local bijection which sends each atomic element of T and each exception of π lying in or mapped into a litter in T to its image under π and fixes each atomic element of $S \setminus S'$. We claim that the B-allowable permutation π' extending this local bijection with no exceptions outside the domain of the local bijection agrees with π 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 π 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 τ^2_A 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 τ_A^2 which essentially has L in its support. But a set in τ_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 τ_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 π might conflict with the need to fix litters in $S \cap \tau_{A_2}^1$.

Thus $\pi(Y) = \pi'(Y)$. Further, π' 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 π' on its support; and exception discipline prevents it from moving the litters themselves because π' has no exceptions in relevant τ_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 π^{-1} shows that $\pi(X) = X$ as desired.

The size of sets τ_A^2 is μ

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

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

We introduce another special kind of support. A nice A-support is a strict A-support in which each atom in the domain either belongs to no near-litter in the domain or is preceded by a near-litter containing it, in which distinct near-litters in the domain are disjoint, and in which each litter L belonging to a τ_B^1, $B \ll A$, is preceded by a B-support of $\Pi_B([L])$. There is a certain general similarity to strong supports, but notice that litters in a nice support do not have to be near-litters, and that the image of a nice A-support under an A-allowable permutation is actually a nice support.

If S is a support of x, we define the coding function $\xi_{x,S}$ so that $\pi(x) = \xi_{x,S}(\pi(S))$ for each A-allowable permutation π. Notice that if $\pi(S) = \pi'(S)$ then $\pi' \circ \pi^{-1}$ fixes S and thus fixes x, so $\pi(x) = \pi'(x)$. The coding function $\xi_{x,S}$ is a bijection from the orbit of S under A-allowable permutations to the orbit of x.

Our strategy is to show that there is a covering set of coding functions of size $< \mu$ for elements of τ_A^2 for each A (that is, a set of coding functions such that every element of $\tau^2(A)$ is a value of a coding function in this set). Since there are μ supports, this establishes the desired result.

We describe a combinatorial object associated with a nice A-support S called its A-format: replace each atom in τ_B^1 in the support with $(\alpha, (1, B, \beta))$ where α is the ordinal position of the item and β is the ordinal position of a near-litter N including the atom if there is one ($\beta < \alpha$ in this case) or κ if there is no such occurrence; replace each near-litter in τ_B^1 with $(\alpha, (2, \beta))$ if α is the position of the litter, $B \ll A_1$, and β is the position of the atom $\Pi(B)([N^\circ])$ in the support, with $(\alpha, (3, \chi))$ if $B \ll A_1$ and $\chi(S_B^L) = L$, where χ is a coding function and S_B^L is the maximal strict B-support included in the segment in S before L, and with $(\alpha, (4, \emptyset))$ if $B = A_1$. We assume that coding functions χ are taken from small covering sets of coding functions assumed to exist by inductive hypothesis. It should be clear that when we set to work on A, we already know that there are $< \mu$ B-formats for each $B \ll A$,

because we know by ind hyp that covering families are small, and otherwise they are structures of cardinality small relative to μ made up of components taken from sets of size suitably small relative to μ.

Formats of A-supports correspond precisely to orbits in A-allowable permutations of supports.

Clearly, if an A-support has a certain format, its image under an A-allowable permutation will have the same format. All the conditions describing the format are invariant under application of an allowable permutation of sufficient index.

Now suppose that supports S and T have the same format. Define a local bijection sending S to T in a natural sense: each atom in S should be mapped to the atom in the corresponding position in T. This has the effect of ensuring that local cardinals of litters in S are mapped to the corresponding litters in T by a permutation extending the local bijection, if all earlier items in S have been successfully mapped to the corresponding items in T, because a support for the local cardinal has been

handled correctly. We do a little extra work to ensure that near-litters in corresponding positions are mapped exactly to each other. If L corresponds to M, we have $[L]$ mapped to $[M]$. Consider a specific object x in $L \setminus L^\circ$. Choose a sequence $x_0, x_1 \ldots$ of items in L° and a sequence x'_0, x'_1 of items in M°, none of these lying in the domain of the local bijection. Augment the local bijection to map x to x'_0 and each x_i to x'_{i+1}. Similarly, if $y \in M \setminus M^\circ$, choose (for each such y) sequences y_i in M° and y'_i in L°, and arrange for y'_0 to be mapped to y and for y'_{i+1} to be mapped to y_i. Each new orbit then needs to be completed in a way which creates no new exceptions in litters in S or T. An A-allowable permutation extending the resulting local bijection and having no exceptions outside its domain sends S exactly to T, so indeed the formats capture the orbits in A-supports determined by A-allowable permutations.

We develop a representation for an $X \in \tau_A^2$. For each element Y of X, choose a small support consisting of atoms in τ_A^0 and litters in τ_A^1. For each litter L in the support such that $\Pi_A([L]) \in \tau_B^2$, choose a B-nice support T such that $\xi_{\Pi_A([L]),T}$ is in the appropriate covering set of coding functions of size $< \mu$. For each other litter in the support, we get $\Pi_A([L])$ an atom in $\tau_{A_1}^0$.

Merge all of this data into a nice support S_Y for Y with S as an initial segment. Notice that we can ensure that $< \mu$ formats for supports S_Y are used: the exact data is a small sequence of atoms (information about which in the format is coded by a very limited range of ordinals) and a small sequence of nice B-supports as above, taken from small covering sets, and a further small sequence of atoms from $\tau_{A_1}^0$, and S. Place S initially, arrange the others in whatever order desired then remove duplicates.

Now we claim that a coding function for X is determined by two items, the support S and