This is an essay on the thought process that led to the “tangled web” component of my NF consistency proof.

This started with the idea of trying to reverse engineer the NFU consistency proof of Jensen in a way which would yield a possible approach to a consistency proof for NF.

Here is a presentation of the Jensen consistency proof.

A natural model of TSTU is obtained from a sequence of sets $X_i$ ($i \in \mathbb{N}$) such that $\mathcal{P}(X_i) \subseteq X_{i+1}$ (or, more abstractly, we have an injection $f_i : \mathcal{P}(X_i) \to X_{i+1}$ for each $i$). We translate a sentence of the language of TSTU into terms of this model by interpreting each quantifier over a type $i$ variable as a quantifier restricted to $X_i$ in the ambient set theory, and interpreting $x^i = y^i$ just as in the ambient theory, and interpreting $x^i \in y^{i+1}$ as $x^i \in y^{i+1} \land y^{i+1} \in \mathcal{P}(X_i)$ in the sense of the ambient theory in the first version proposed, or as $(\exists y \in \mathcal{P}(X_i) : x^i \in y \land f_i(y) = y^{i+1})$ in the more abstract version.

For our purposes, we will use the more concrete version, stipulating that each $X_i$ is a transitive set, and $\mathcal{P}(X_i) \subseteq X_{i+1}$ for each $i$. Our approach can be adapted to the more abstract definition of natural model, it should be noted.

Our conditions ensure that not only is $\mathcal{P}(X_i)$ a subset of $X_{i+1}$, but in fact it is a subset of each $X_j$ for $j > i$. This has the effect that we can skip types: any strictly increasing sequence $s_i$ of natural numbers determines an interpretation of TSTU in which type $i$ is implemented as $X_{s_i}$ and $x^i \in y^{i+1}$ is interpreted as $x^i \in y \land f_i(y) = y^{i+1}$.

For generality, note that we can provide types $X_\alpha$ for each $\alpha \in \lambda$, a fixed limit ordinal, with each $X_\alpha$ transitive and $\mathcal{P}(X_\alpha) \subseteq X_\beta$ for each $\beta > \alpha$, and an increasing sequence $s$ of elements of $\lambda$ will determine an interpretation of TSTU just as above.

It is because we can skip types in this way that the Ramsey theorem becomes applicable. A finite set $\Sigma$ of formulas of the language of TSTU will determine a partition of $|\lambda|^n$, where $n$ is a strict upper bound on the type indices appearing in $\Sigma$, in which each $A \in [\lambda]^n$ will go into a compartment determined by the truth values of the elements of $\Sigma$ in models of TSTU determined as above by sequences $s$ such that the range of the restriction of $s$ to $\{0, 1, 2, \ldots, n-1\}$ is $A$. This partition has no more than $2^{2|\Sigma|}$ compartments. By the Ramsey theorem it has an infinite homogeneous set $H$. The model of TSTU determined by a
sequence \( s \) with range included in \( H \) will satisfy Ambiguity for each formula in \( \Sigma \). Thus the full ambiguity scheme is consistent with TSTU by compactness, and NFU is consistent by Specker’s results.

Adapting this to NF would require the ability to skip types as we do in the argument above. This seems a very odd thing to do, but it is possible to describe situations which are not obviously impossible in which it could be done.

The first theory of this kind that I defined is tangled type theory (TTT). This is the first-order theory with equality and membership, with sorts indexed by elements of a limit ordinal \( \lambda \), in which a sentence \( x = y \) is well formed iff the ordinal index of the sort of \( x \) is equal to the ordinal index of the sort of \( y \), and a sentence \( x \in y \) is well-formed iff the ordinal index of the sort of \( x \) is less than the ordinal index of the sort of \( y \). For each formula of the language of TST and each strictly increasing sequence \( s \) with range in \( \lambda \), the formula \( \phi^s \) is the formula of the language of TTT obtained by replacing each type \( i \) variable in \( \phi \) with a type \( s_i \) variable. This gives a well-formed formula of TTT. The axioms of TTT are exactly the formulas \( \phi^s \) where \( s \) is any strictly increasing sequence in \( \lambda \) and \( \phi \) is an axiom of TST (recalling that these are axioms of extensionality and comprehension).

TTT is consistent if and only if NF is consistent. If NF is consistent, we readily obtain a model of TTT by using a copy of our model of NF to implement every type of TTT, and defining the membership relation between any two types using the membership relation of the model of NF in the obvious way.

Note a model of TTT and any strictly increasing sequence \( s \) in \( \lambda \) determine a model of TST in which type \( i \) is implemented as type \( s_i \) of the model of TTT and membership of type \( i \) in type \( i + 1 \) in the model of TST is determined by membership of type \( s_i \) in type \( s_{i+1} \) in the model of TTT.

If TTT is consistent, we follow the Jensen proof. Once again, it is because we can “skip types” that the Ramsey theorem becomes applicable. A finite set \( \Sigma \) of formulas of the language of TST will determine a partition of \( [\lambda]^n \), where \( n \) is a strict upper bound on the type indices appearing in \( \Sigma \), in which each \( A \in [\lambda]^n \) will go into a compartment determined by the truth values of the elements of \( \Sigma \) in models of TST determined as above by a given model of TTT and any strictly increasing sequence \( s \) with range included in \( \lambda \) such that the range of the restriction of \( s \) to \( \{0, 1, 2, \ldots, n-1\} \) is \( A \). This partition has no more than \( 2^{2^n} \) compartments. By the Ramsey theorem it has an infinite homogeneous set \( H \). The model of TST determined by the given model of TTT and a strictly increasing sequence \( s \) with range included in \( H \) will satisfy Ambiguity for each formula in \( \Sigma \). Thus the full ambiguity scheme is consistent with TST by compactness, and NF is consistent by Specker’s results.

Now TTT looks perfectly mad. There can be no natural models of TTT, since one cannot arrange for each type to be identified with, or even to be the same size as, the power set of each lower type, which is what the internal language of the theory suggests is intended.

We “unfold” the type system of TTT to get the notion of a tangled web.

Observe first that we want to use a more abstract notion of what a natural model of TST is in this context. A natural model is determined by a sequence
$X_i$ of sets ($i \in \mathbb{N}$) in which there is a bijection $f_i : X_{i+1} \to \mathcal{P}(X_i)$. A formula of TST can be translated into terms of a natural model by replacing quantifiers over type $i$ with quantifiers restricted to $X_i$, reading $x^i = y^i$ in the ambient set theory as equality, and reading $x^i \in y^{i+1}$ as $x^i \in f_i(y^{i+1})$. We will also talk about models of TST, the restriction of TST to $n$ types.

The theory of a natural model of TST (or of TST$_n$) is entirely determined by the cardinality of its base type $X_0$.

There is an internal notion of a natural model in TST or TTT. A natural model is determined by a sequence of sets (possibly finite) $X_i$ where there is a bijection from $i^*X_{i+1}$ to $\mathcal{P}(X_i)$ for each $i$ ($i$ being the singleton operation). The definition goes as above with $x^i \in y^{i+1}$ being interpreted as $x^i \in f_i(y^{i+1})$.

In TST, let $V^i$ be the universal set of type $i$. In TST, any finite sequence $X_i = \iota_{m-i} V_{k+i}$ with $f_i$ mapping an element $\iota^{m-i}_j(A)$ of $X_{i+1}$ to

$$\{\iota^{m-i}_j(a) : a \in A\} \in \mathcal{P}(X_i),$$

determines a natural model of a TST$_n$ (the number of types has not been specified). Notice that here all the $X_i$'s are sets of the same type $m + k$, being images of lower types elementwise under a sufficient number of iterations of the singleton map to bring them up to the target type.

This can be emulated in TTT and gives a considerably more complicated picture. For any type indices $i < j$, let $\iota_{j,i}$ be the singleton map from type $i$ objects to type $j$ objects. Obviously a sequence of one set $V_j$ (the universe of type $i$) determines a natural model of TST$_1$. If we have a finite sequence $\tau$ whose final element is $V_j$ determining a natural model of TST$_k$, with $j < i$, then the sequence $\tau^i$ with $\tau^i_n = \iota_{j,i} \tau_n$ for $n < k$ and $\tau^i_k = V_j$ determines a natural model. In this way we can get a natural model determined by any finite increasing sequence of types. For each finite set $A$ of type indices, we define a cardinal $\tau_A$ by an annoying induction. $\tau(\{i\}) = |V_i|$. If $A$ has $\alpha < \beta$ as its two largest elements and $\tau(A \setminus \{\beta\}) = |X|$ then $\tau(A) = \iota_{\alpha,\beta}X$. We can establish that all $\tau(A)$ with $A$ having the same maximum element are of the same type, and if $A$ has more than one element, $\exp(\tau(A)) = \tau(A \setminus \{\min(A)\})$ [exp being the map $|X| \mapsto \mathcal{P}(X)$], and further that for any concrete sentence $\phi$, the truth value of $\phi$ in the theory of the model of TST$_n$ with base type $\tau(A)$ is determined by the first $n$ elements of $A$ (because this is an internalization of the truth value of $\phi$ in the part of the model of TTT determined by those types). There is no reason to believe that the limited elementary equivalence just described can be internalized, because internally we may have nonstandard "sentences" to consider.

This motivates a procedure to be carried out in untyped set theory which we now describe.

Provide cardinals indexed not by elements of $\lambda$ but by finite subsets of $\lambda$. The cardinal indexed by a nonempty subset $A$ of $\lambda$ we denote by $\tau(A)$. A set of cardinality $\tau(A)$ is to be thought of as one of many pictures of type $\min(A)$ in a model of TTT, though this plays no direct role in our definition.

We define $A_1$ as $A \setminus \{\min(A)\}$. The power set of a set of size $\tau(A)$ is to be of size $\tau(A_1)$, if $A$ has at least two elements [this is our device for type-skipping;
this set thought of as a copy of the type indexed by the smallest element of $A$. The theory of the first $n$ types of a natural model of TST with base type of size $\tau(A)$ should depend only on the first $n$ elements of $A$, if $|A| \geq n$: the idea is that when we label the types of a natural model with cardinals $\tau(A)$, the theory of any finite interval of types in such a model is determined entirely by the minima of the sets $A$, following our idea that a set of size $\tau(A)$ is a sort of copy of type $\min(A)$ in an imagined model of TTT.

Leaving murky motivations aside, we give the formal definition:

A tangled web of cardinals is a function $\tau$ from nonempty finite subsets of $\lambda$ [a limit ordinal given in the context] satisfying two conditions:

**naturality**: $2^{\tau(A)} = \tau(A_1)$, if $|A| \geq 2$.

**elementarity**: The first-order theory of a natural model with type 0 implemented by a set of size $\tau(A)$ depends only on the $n$ smallest elements of $A$, if $|A| \geq n$. For technical reasons having to do with our actual construction in the NF consistency proof, we actually modify this to “The first-order theory of natural models of $\text{TST}_n$ with type 0 implemented by a set of size $\tau(A)$ depends only on the $n + 1$ smallest elements of $A$, if $|A| \geq n + 1$.”

The existence of a tangled web implies the consistency of NF by a direct adaptation of Jensen’s argument:

A finite set $\Sigma$ of formulas of the language of TST will determine a partition of $[\lambda]^n$, where $n$ is a strict upper bound on the type indices appearing in $\Sigma$, in which each $A \in [\lambda]^n$ will go into a compartment determined by the truth values of the elements of $\Sigma$ in natural models of $\text{TST}_n$ with base type $\tau(A)$. This partition has no more than $2^{|\Sigma|}$ compartments. By the Ramsey theorem it has a homogeneous set $H$ of size $n + 2$. Any natural model of $\text{TST}_{n+1}$ with base type of size $\tau(H)$ will satisfy Ambiguity for every formula in $\Sigma$, whence Ambiguity for formulas in $\Sigma$ is consistent with TST, so the full Ambiguity scheme is consistent with TST by compactness, whence NF is consistent by results of Specker.

It should be noted that in the structures internal to TTT one does not actually find a tangled web. As sketched above, one can internally define a sort of tangled web with $\lambda$ a concrete natural number: the naturality condition will hold, but the elementarity condition will hold for each concretely given sentence rather than for the entire first-order theory (nonstandard sentences may behave badly).