New Foundations is consistent

Sky Wilshaw

July 2023

Abstract

We give a self-contained account of a version of Holmes' proof [2] that Quine's set theory *New Foundations* [4] is consistent relative to the metatheory ZFC. This is a 'deformalisation' of the formal proof written in Lean at [7].

Contents

Overview

In §1, we outline the context for the proof we will present. The mathematical background expected in subsequent sections will be limited to basic familiarity with cardinals and ordinals. We will then give an outline of the proof in §2. In §3, we introduce some basic preliminaries, and explicitly describe the structure within which our model will reside.

All proofs given in §3 are verified by the theorem prover Lean.

Note: At the present time, the formal proof [7] is incomplete, and this paper reflects the unfinished state of that proof. We aim to keep this paper proof in line with the formal proof, although as the project is ongoing, some variance is to be expected. The current version of the paper is available at `https://zeramorphic.github.io/con-nf-paper/main.pdf`.

1 The theories at issue

In 1937, Quine introduced *New Foundations* (NF) [4], a set theory with a very small collection of axioms. To give a proper exposition of the theory that we intend to prove consistent, we will first make a digression to introduce the related theory TST, as explained by Holmes in [2]. We will then describe the theory TTT, which we will use to prove our theorem.

1.1 The simple theory of types

The *simple theory of types* (known as *théorie simple des types* or TST) is a first order set theory with several sorts, indexed by the nonnegative integers. Each sort, called a *type*, is comprised of *sets* of that type; each variable x has a nonnegative integer $\texttt{type}(\text{`}x\text{'})$ which denotes the type it belongs to. For convenience, we may write x^n to denote a variable x with type n.

The primitive predicates of this theory are equality and membership. An equality '$x = y$' is a well-formed formula precisely when $\texttt{type}(\text{`}x\text{'}) = \texttt{type}(\text{`}y\text{'})$, and similarly a membership formula '$x \in y$' is well-formed precisely when $\texttt{type}(\text{`}x\text{'}) + 1 = \texttt{type}(\text{`}y\text{'})$.

The axioms of this theory are extensionality

$$\forall x^{n+1}.\, \forall y^{n+1}.\, (\forall z^n.\, z^n \in x^{n+1} \leftrightarrow z^n \in y^{n+1}) \to x^{n+1} = y^{n+1}$$

and comprehension

$$\exists x^{n+1}.\, \forall y^n.\, (y^n \in x^{n+1} \leftrightarrow \varphi(y^n))$$

where φ is any well-formed formula, possibly with parameters.

Remarks 1.1. (i) These are both axiom schemes, quantifying over all type levels n, and (in the latter case) over all well-formed formulae φ.

(ii) The inhabitants of type 0, called *individuals*, cannot be examined using these axioms.

(iii) By comprehension, there is a set at each type that contains all sets of the previous type. Russell-style paradoxes are avoided as formulae of the form $x^n \in x^n$ are ill-formed.

1.2 New Foundations

New Foundations is a one-sorted first-order theory based on TST. Its primitive propositions are equality and membership. There are no well-formedness constraints on these primitive propositions.

Its axioms are precisely the axioms of TST with all type annotations erased. That is, it has an axiom of extensionality

$$\forall x.\, \forall y.\, (\forall z.\, z \in x \leftrightarrow z \in y) \to x = y$$

and an axiom scheme of comprehension

$$\exists x.\, \forall y.\, (y \in x \leftrightarrow \varphi(y))$$

the latter of which is defined for those formulae φ that can be obtained by erasing the type annotations of a well-formed formula of TST. Such formulae are called *stratified*. To avoid the explicit dependence on TST, we can equivalently characterise the stratified formulae as follows. A formula φ is said to be stratified when there is a function σ from the set of variables to the nonnegative integers, in such a way that for each subformula '$x = y$' of φ we have $\sigma(\text{`}x\text{'}) = \sigma(\text{`}y\text{'})$, and for each subformula '$x \in y$' we have $\sigma(\text{`}x\text{'}) + 1 = \sigma(\text{`}y\text{'})$.

Remarks 1.2. (i) It is important to emphasise that while the axioms come from a many-sorted theory, NF is not one; it well-formed to ask if any set is a member of, or equal to, any other.

(ii) Russell's paradox is avoided because the set $\{x \mid x \notin x\}$ cannot be formed; indeed, $x \notin x$ is an unstratified formula. Note, however, that the set $\{x \mid x = x\}$ is well-formed, and so we have a universe set.

(iii) The infinite set of stratified comprehension axioms can be described with a finite set; this is a result of Hailperin [1].

(iv) Specker showed in [5] that NF disproves the Axiom of Choice.

While our main result is that New Foundations is consistent, we attack the problem by means of an indirection through a third theory.

1.3 Tangled type theory

Introduced by Holmes in [3], *tangled type theory* (TTT) is a multi-sorted first order theory based on TST. This theory is parametrised by a limit ordinal λ, the elements of which will index the sorts. As in TST, each variable x has a type that it belongs to, denoted $\texttt{type}(\text{`}x\text{'})$. However, in TTT, this is not a positive integer, but an element of λ.

The primitive predicates of this theory are equality and membership. An equality '$x = y$' is a well-formed formula when $\texttt{type}(\text{`}x\text{'}) = \texttt{type}(\text{`}y\text{'})$. A membership formula '$x \in y$' is well-formed when $\texttt{type}(\text{`}x\text{'}) < \texttt{type}(\text{`}y\text{'})$.

The axioms of TTT are obtained by taking the axioms of TST and replacing all type indices in a consistent way with elements of λ. More precisely, for any order-embedding $s : \omega \to \lambda$, we can convert a well-formed formula φ of TST into a well-formed formula φ^s of TTT by replacing a type variable α with $s(\alpha)$.

Remarks 1.3. (i) Membership across types in TTT behaves in some quite bizarre ways. Let $\alpha \in \lambda$, and let x be a set of type α. For any $\beta < \alpha$, the extensionality axiom implies that x is uniquely determined by its type-β elements. However, it is simultaneously determined by its type-γ elements for any $\gamma < \alpha$. In this way, one extension of a set controls all of the other extensions.

(ii) The comprehension axiom allows a set to be built which has a specified extension in a single type. The elements not of this type may be considered 'controlled junk'.

We now present the following striking theorem.

Theorem 1.4 (Holmes)**.** NF is consistent if and only if TTT is consistent.

The proof is not long, but is outside the scope of this paper; it requires more model theory than the rest of this paper expects a reader to be familiar with, and relies on additional results such as those proven by Specker in [6].

Thus, our task of proving NF consistent is reduced to the task of proving TTT consistent. We will do this by exhibiting an explicit model (albeit one that requires a great deal of Choice to construct). As TTT has types indexed by a limit ordinal, and sets can only contain sets of lower type, we can construct a model by recursion over λ. This was not an option with NF directly, as the universe set $\{x \mid x = x\}$ would necessarily be constructed before many of its elements.

2 Outline

To construct a model of tangled type theory, we build each type individually, and then prove that the resulting structure satisfies the required axioms. The process for building each type is complicated, and depends on some knowledge about the construction of the previous types. In the following subsections, we outline the construction the types, as well as the precise facts we need to carry through the inductive hypothesis at each stage.

2.1 Model parameters

As described in §1.3, the types of a given model of tangled type theory are indexed by a limit ordinal λ. Our model will also have two more cardinal parameters, denoted κ and μ, satisfying $\lambda < \kappa < \mu$.

Sets smaller than size κ will be called *small*. We require that κ is a regular cardinal; this ensures that small-indexed unions of small sets are small.

Each type in our model will have size μ. We require μ to be a strong limit cardinal; power sets of sets smaller than μ must also be smaller than μ. We stipulate that the cofinality of μ is at least κ. This assumption will become important whenever we consider objects indexed by small ordinals.

We remark that these constraints are satisfiable; $\lambda = \aleph_0, \kappa = \aleph_1, \mu = \beth_{\omega_1}$ suffice.

2.2 Atoms and permutations

To aid our construction, we will add an additional level of objects below type zero. These will not be a part of the final model we construct. This base type will be comprised of objects called *atoms* (although they are not atoms in the traditional model-theoretic sense).

Alongside the construction of the types of our model, we will also construct a collection of permutations of each type, called the *allowable permutations*. Such permutations will preserve the structure of the model in a strong sense; for instance, they preserve membership.

2.3 Construction of each type

Objects in our model are defined by their elements at all lower type indices. However, not all collections of extensions may become model elements; for example, they may fail to satisfy extensionality at all levels simultaneously. We impose two restrictions on what kind of extensions an object may have.

The first restriction is that one of the extensions of a given object must be 'preferred', and every other extension must be easily derivable from that particular extension. This will help us to establish extensionality, as model elements will be the same if and only if their preferred extensions are the same. The system to compute other extensions uses a construction called the *fuzz* map. This map turns information about one extension into 'ordered junk' in another extension, in such a way that the model cannot learn anything useful about the non-preferred extensions. Our allowable permutations will be defined as a set of permutations that respect the fuzz map.

The second restriction is that the object must have a small *support* comprised of *addresses*. That is, the behaviour of the object under the action of allowable permutations must be fully characterisable by the behaviour of a small set of addresses under allowable permutations. This will ensure that the objects of our model are not too complex. Because the cofinality of μ is at least κ, there are only

μ small sets of elements taken from a collection of size μ; this observation will play a key role in establishing the sizes of our types.

2.4 Constraining the size of each type

The construction of a given type can only be done under the assumption that each smaller type was of size exactly μ. This means that we need to prove that each type has size μ in the inductive step. In order to do this, we will need to show that there are a lot of allowable permutations. The main theorem establishing this, called the *freedom of action theorem*, roughly states that under certain assumptions, a permutation defined on a small set of addresses can be extended to an allowable permutation. The majority of this paper will be allocated to proving the freedom of action theorem, and it will be outlined in more detail when we are in a position to prove it. One this is established, we can prove that the size of each type is precisely μ by carefully counting the possible ways to describe a model element.

2.5 Finishing the induction

We can then finish the inductive step and build the entire model. It remains to show that this is a model of TTT as desired. This part of the proof is quite direct, and also uses the freedom of action theorem.

3 The underlying structure

3.1 Atoms, litters, and near-litters

As described in §2.2, we have an additional level of objects below type zero. To index the levels of the model, together with this new level, we make the following definition.

Definition 3.1. A *type index* is an element of λ or a distinguished symbol $\bot$. We impose an order on type indices by setting $\bot < \alpha$ for all $\alpha \in \lambda$. The set of type indices is denoted $\lambda^\bot$.

Elements of λ may be called *proper type indices*.

Our base type is a set of *atoms*, organised into *litters*.

Definition 3.2. A *litter* is a triple $L = (\nu, \beta, \gamma)$ where $\nu \in \mu$, β is a type index, and $\gamma \neq \beta$ is a proper type index.

This somewhat arcane definition will be used to great effect later when defining the fuzz map. A litter $L = (\nu, \beta, \gamma)$ encodes data coming from type β and going into type γ. Note that β may be $\bot$, but γ may not; this corresponds to the fact that we never construct data in type $\bot$ from data at higher levels. The first component ν is an index allowing us to have μ distinct litters with the same source and target types.

Remark 3.3. There are precisely μ litters.

Definition 3.4. An *atom* is a pair $a = (L, i)$ where L is a litter and $i \in \kappa$. The *associated litter* of an atom is its first projection $\mathsf{pr}_1(a)$, written a° for brevity. The *litter set* $\mathsf{LS}(L)$ of a given litter L is the set of atoms whose associated litter is L; that is, $\mathsf{LS}(L) = \{(L, i) \mid i \in \kappa\}$. The litter sets partition the set of atoms into μ sets of κ atoms.

Remark 3.5. Many of our constructions rely on having only a small set of constraints. If our constraints take the form of atoms, the smallness assumption guarantees that most of the atoms in a given litter are unconstrained. Motivated by smallness concerns, we make the following definition.

Definition 3.6. A *near-litter* is a pair $N = (L, s)$ where L is a litter and s is a set of atoms with small symmetric difference to the litter set of L. We say that the *associated litter* of N is $N^\circ = \mathrm{pr}_1(N)$, or that N is *near* L.

Remarks 3.7. (i) A set of atoms can be near at most one litter. For brevity, we will frequently identify a near-litter with its underlying set.

(ii) The litter set of any litter L can be made into a near-litter: $(L, \mathsf{LS}(L))$.

(iii) Each near-litter has size exactly κ, and there are μ near-litters in total; the latter follows from the fact that the cofinality of μ is at least κ.

We can now define the allowable permutations of type $\bot$, although we will give them a different name for now; they will be precisely those permutations of atoms that respect the structure of near-litters.

Definition 3.8. A *near-litter permutation* is a permutation π of atoms that sends near-litters to near-litters.

Remarks 3.9. (i) A near-litter permutation π induces a permutation of litters, which we will also call π. This is defined by mapping L to the associated litter of $\pi''\mathsf{LS}(L)$, where the double apostrophe denotes pointwise function application ($f''s$ denotes the set $\{f(x) \mid x \in s\}$). Thus, a near-litter permutation is simultaneously a permutation of atoms, litters, and near-litters.

(ii) The set of near-litter permutations forms a group under composition.

References

[1] Theodore Hailperin. "A set of axioms for logic". In: *Journal of Symbolic Logic* 9.1 (1944), pp. 1–19. DOI: 10.2307/2267307.

[2] M. Randall Holmes. *NF is Consistent*. 2023. arXiv: 1503.01406 [math.LO].

[3] M. Randall Holmes. "The Equivalence of NF-Style Set Theories with "Tangled" Theories; The Construction of ω-Models of Predicative NF (and more)". In: *The Journal of Symbolic Logic* 60.1 (1995), pp. 178–190. ISSN: 00224812. URL: http://www.jstor.org/stable/2275515.

[4] W. V. Quine. "New Foundations for Mathematical Logic". In: *American Mathematical Monthly* 44 (1937), pp. 70–80. URL: https://api.semanticscholar.org/CorpusID:123927264.

[5] Ernst P. Specker. "The Axiom of Choice in Quine's New Foundations for Mathematical Logic". In: *Proceedings of the National Academy of Sciences of the United States of America* 39.9 (1953), pp. 972–975. ISSN: 00278424. URL: http://www.jstor.org/stable/88561.

[6] Ernst P. Specker. "Typical Ambiguity". In: *Logic, Methodology and Philosophy of Science*. Ed. by Ernst Nagel. Stanford University Press, 1962, pp. 116–123.

[7] Sky Wilshaw et al. *The consistency of New Foundations*. 2022–2023. URL: https://leanprover-community.github.io/con-nf/.