New Foundations is consistent

Sky Wilshaw

July 2023

Abstract

We give a self-contained account of a version of Holmes' proof [4] that Quine's set theory *New Foundations* [7] is consistent relative to the metatheory ZFC. This is a 'deformalisation' of the formal proof written in Lean at [10].

Contents

1   Overview

In §2, 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 §3. In §4, we introduce some basic preliminaries, and explicitly describe the structure within which our model will reside. The objects of our model will be constructed in §5. We then show the main technical theorem, the *freedom of action theorem*, in §6.

All proofs given in sections 4 to 6 are verified by Lean [6], which is a functional programing language and interactive theorem prover. Most of our proofs are written in Lean's *tactic mode*, in which Lean tracks a number of hypotheses and goals, and we instruct it via tactics to update these hypotheses and goals according to logical rules. There are a large variety of tactics to perform different tasks, such as simplification (`simp`), rewriting of subexpressions (`rw`), case splitting (`cases` and `by_cases`), and so on. These tactics output a *proof term*, which is then passed to Lean's *kernel*. The kernel is a small program that verifies proofs by performing a type-checking algorithm. Lean's kernel is the only code that must be 'trusted' in order to guarantee that a proof accepted by Lean as a whole is correct, since any unsound tactics would output a proof term that is not type-correct.

We made frequent use of the community-made repository `mathlib` [2], which encodes standard mathematical definitions and theorems in Lean; without this, we would have needed to write our own libraries for abstract algebra and cardinal and ordinal arithmetic.

It is important to note that although theorems can be verified by Lean's kernel, it has no way of knowing whether a formal definition written in Lean matches the familiar mathematical definition; any definitions used in a theorem statement must be manually checked. For example, if verifying a proof of Fermat's last theorem, it would suffice to manually check the definitions of natural numbers, addition, exponentiation, and natural equalities and inequalities.

Carneiro proved in [1] that Lean's type theory is consistent relative to

$$\text{ZFC} + \{\text{there are } n \text{ inaccessible cardinals} \mid n < \omega\}$$

These inaccessible cardinals are needed to support Lean's hierarchy of type universes. Higher universes are used by `mathlib` whenever they are convenient, for example in its definitions of cardinals and ordinals. However, these uses are not strictly necessary for our purposes, and the entire proof can be carried out in plain ZFC, as shown by [4] and this paper.

Note: At the present time, the formal proof [10] 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.

2   The theories at issue

In 1937, Quine introduced *New Foundations* (NF) [7], 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 [4]. We will then describe the theory TTT, which we will use to prove our theorem.

2.1 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 $\mathsf{type}('x')$ 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 $\mathsf{type}('x') = \mathsf{type}('y')$, and similarly a membership formula '$x \in y$' is well-formed precisely when $\mathsf{type}('x') + 1 = \mathsf{type}('y')$.

The axioms of this theory are extensionality

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

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 2.1.   (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.

2.2 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

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

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 $\sigma('x') = \sigma('y')$, and for each subformula '$x \in y$' we have $\sigma('x') + 1 = \sigma('y')$.

Remarks 2.2.   (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 [3].

(iv) Specker showed in [8] 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.

2.3 Tangled type theory

Introduced by Holmes in [5], *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 $\texttt{type}(`x')$. 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 $\texttt{type}(`x') = \texttt{type}(`y')$. A membership formula '$x \in y$' is well-formed when $\texttt{type}(`x') < \texttt{type}(`y')$.

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 2.3.   (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 2.4 (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 [9].

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.

3 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.

3.1 Model parameters

As described in §2.3, 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.

3.2 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.

3.3 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.

3.4 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.

3.5 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.

4 The underlying structure

4.1 Atoms, litters, and near-litters

As described in §3.2, 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 4.1. 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 4.2. 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 4.3. There are precisely μ litters.

Definition 4.4. 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 $\mathrm{pr}_1(a)$, 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, and there are μ atoms in total.

Remark 4.5. 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 4.6. 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 $N° = \mathsf{pr}_1(N)$, or that N is *near* L.

Remarks 4.7. (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: $\mathsf{NL}(L) = (L, \mathsf{LS}(L))$.

(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 $\perp$, 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 4.8. A *near-litter permutation* π is a permutation of atoms that sends near-litters to near-litters.

Remarks 4.9. (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.

4.2 Higher type structure

A type-α object has elements of any type $\beta < \alpha$, which have elements of any type $\gamma < \beta$, and so on; we must eventually reach $\perp$ in a finite number of steps by well-foundedness. We will now make a definition to deal with sequences of type indices obtained in this way.

Definition 4.10. A *path of type indices* $\alpha \rightsquigarrow \varepsilon$ is a finite sequence of type indices

$$\alpha > \beta > \gamma > \cdots > \varepsilon$$

If α is a type index, an *α-extended type index* is a path $\alpha \rightsquigarrow \perp$. If A is a path $\alpha \rightsquigarrow \beta$ and B is a path $\beta \rightsquigarrow \gamma$, their composition $A \gg B$ is a path $\alpha \rightsquigarrow \gamma$ obtained by concatenation of the sequences but removing the duplicated index β.

Remark 4.11. For any $\alpha \in \lambda^\perp$, the set of α-extended type indices has size at most λ.

In our model, the iterated extensions of objects of type α are indexed by the α-extended type indices. We can apply operations to an object along each path $\alpha \rightsquigarrow \perp$. One important notion defined in this way will be *structural permutations*.

Definition 4.12. An *α-structural permutation* π assigns a near-litter permutation to each α-extended index. If A is an α-extended index, the assigned near-litter permutation is called the *derivative* of π along A, and is denoted π_A. We can extend this notion to arbitrary paths $\alpha \rightsquigarrow \beta$; in this case, the derivative π_A is a β-structural permutation. The group of α-structural permutations is denoted S_α.

Remark 4.13. Near-litter permutations may be identified with $\perp$-structural permutations.

To apply an α-structural permutation π to some model element x of type α, we first consider its elements of some type $\beta < \alpha$. By recursion, we apply $\pi_{(\alpha > \beta)}$ to each such element, where $\pi_{(\alpha > \beta)}$ is the derivative of π along the one-step path from α to β. These images are then assembled to form a new object of type α; note however that this new object need not in general satisfy the constraints

outlined in §3.3, so structural permutations need not map elements of the model to other elements of the model.

4.3   Addresses and supports

We are interested in objects that can be characterised by a small amount of data; this data will take the form of *addresses*.

Definition 4.14. Let α be a type index. An α-*address* is a pair (A, x) where A is an α-extended type index and x is either an atom or a near-litter.

An α-address encodes a set of atoms, and a path to get there from a type-α object by descending through iterated extensions.

Note. In [4], addresses are called *support conditions*.

Remark 4.15. The set of α-addresses has cardinality μ. This follows from remarks 4.7 and remark 4.11.

The group of α-structural permutations acts on the set of α-addresses by

$$\pi(A, x) = (A, \pi_A(x))$$

We would like to say that a *support* is a small set of α-addresses; for technical reasons that will be discussed later it will actually be better to make the following definition instead.

Definition 4.16. An α-support is a function S whose domain is a small ordinal and whose values are α-addresses, such that whenever N_1, N_2 are near-litters in the range of S, the atoms in their symmetric difference $N_1 \triangle N_2$ are also in the range of S. The *underlying set* of a support is its range.

Remarks 4.17.   (i) An address may occur multiple times in the range of a support, but all we usually care about is the presence or absence of an address. The fact that these addresses have indices in a given support will be useful for some constructions later. The symmetric difference condition is also a technical condition that we will use later, it avoids a problem where initial conditions for a construction could be overspecified.

 (ii) For any small set of addresses s, there is a support S whose range is s together with the small set of atom addresses required to satisfy the symmetric difference constraint.

As with many parts of our construction, we will need to count precisely how many supports there are. We will first need to establish the following lemma.

Lemma 4.18. Let μ be a strong limit cardinal. Then for any $\kappa > 0$ strictly smaller than the cofinality of μ, we have

$$\mu^\kappa = \mu$$

Recall that *König's theorem* states that for an infinite cardinal μ, we have $\mu^{\mathrm{cf}(\kappa)} > \mu$; this lemma is a partial converse to this theorem in the case where μ is a strong limit.

Proof. Let A be a set of size κ, and let B be a set of size μ. We assume that B has a well-ordering, perhaps obtained from the order on μ. Given any function $f : A \to B$, we define a relation $\prec_f$ on A by the inverse image of the ordering on B under f:

$$a \prec_f b \iff f(a) < f(b)$$

One can show by induction on B that any for two functions $f, g : A \to B$, if $\mathrm{im}\, f = \mathrm{im}\, g$ and $\prec_f = \prec_g$, then $f = g$. The amount of pairs $(\mathrm{im}\, f, \prec_f)$ is bounded by μ. Indeed, the amount of possible images

of a function $f : A \to B$ is bounded by μ as κ is less than the cofinality of μ and μ is a strong limit, and the amount of relations $\prec_f$ is bounded by $(2^\kappa)^\kappa$, which is bounded by μ as it is a strong limit. Thus, $\mu^\kappa \leq \mu$, and the converse is clear. □

We can now prove that there are exactly μ supports for any type index α.

Proposition 4.19. Let α be a type index. The set of α-supports has cardinality μ.

Proof. A support is a function from a small ordinal to the set of α-addresses, which has cardinality μ by remark 4.15. Thus, the amount of supports is bounded by

$$\sum_{i<\kappa} \mu^i$$

By lemma 4.18, each summand is precisely μ, and so the sum is precisely $\kappa \cdot \mu = \mu$. □

The α-structural permutations act on α-supports pointwise:

$$\pi(S)(i) = \pi(S(i))$$

We will need to concatenate supports together. The naïve concatenation of supports S_1, S_2 given by

$$S(i) = \begin{cases} S_1(i) & \text{if } i \in \operatorname{dom} S_1 \\ S_2(i - \operatorname{dom} S_1) & \text{otherwise} \end{cases}$$

need not satisfy the symmetric difference condition, so S may not be a support. There is not a unique way to expand this concatenation into a support, so we will instead make the following definition.

Definition 4.20. Let f be a function from a small ordinal whose values are α-support conditions. We say that a support S is a *completion* of f if $S|_{\operatorname{dom} f} = f$ and every address in the extension $S|_{\operatorname{dom} S \setminus \operatorname{dom} f}$ is an atom address (A, x) where $x \in N_1 \triangle N_2$ and $(A, N_1), (A, N_2) \in \operatorname{ran} f$. Every such function has a completion, as the amount of atom addresses in question is small. A support S is a *sum* of supports S_1, S_2 if it is a completion of the concatenation of S_1 and S_2. Every pair of supports has a sum.

5 Constructing the types

In this section, we describe the way in which we construct each type in the model, under the assumption that all previous levels were constructed properly.

5.1 Hypotheses

We explicitly describe the hypotheses required for the inductive step to succeed. The data constructed at each level is called *tangle data*; for uniformity we will also define tangle data at level $\perp$.

Definition 5.1. Let α be a type index. Then *tangle data* at level α consists of

  (i) A set τ_α, called the set of α-*tangles*. Tangles encapsulate our model elements.

 (ii) A group A_α, called the set of α-*allowable permutations*, which acts on τ_α. Allowable permutations will be denoted using the symbol ρ.

(iii) A group homomorphism $A_\alpha \to S_\alpha$, where S_α is the group of α-structural permutations as defined in definition 4.12. We will notationally suppress this homomorphism, and treat A_α as a subgroup of S_α, so allowable permutations act on anything that structural permutations do.

(iv) A function supp that assigns an α-support to each α-tangle. This support must actually support the input tangle: for any α-allowable permutation ρ, if ρ fixes supp t pointwise, it fixes t.

$$(\forall i. \rho((\text{supp } t)(i)) = (\text{supp } t)(i)) \to \rho(t) = t$$

We require that supp commutes with α-allowable permutations ρ in the sense that

$$\text{supp}(\rho(t)) = \rho(\text{supp } t)$$

The tangle data at level $\bot$ is constructed by taking the tangles to be the atoms, the allowable permutations to be the near-litter permutations, and the support function to be given by

$$\text{dom}(\text{supp } a) = 1; \quad (\text{supp } a)(0) = (\varnothing, a)$$

where $\varnothing$ is the empty path $\bot \rightsquigarrow \bot$.

At higher levels $\alpha \in \lambda$, the tangles will be pairs (x, S) where x is a model element of type α and S is a support that supports x under the action of α-allowable permutations. There are μ tangles that encapsulate any given model element. Allowable permutations act on these pairs pointwise.

$$\rho((x, S)) = (\rho(x), \rho(S))$$

It is clear to see that the second projection pr_2 satisfies the requirements of supp defined in (iv) above are satisfied.

Definition 5.2. Let α be a type index with tangle data. Then a *position function* at level α is an injection $\iota_\alpha : \tau_\alpha \to \mu$. For convenience, we will simply write ι for all position functions.

At level $\bot$, the position function is chosen arbitrarily; one exists as there are exactly μ atoms. At higher levels $\alpha \in \lambda$, the existence of a position function depends on the fact that τ_α has size at most μ. We will typically assume that all type levels previously constructed have a position function.

Let $\alpha \in \lambda$, and let N be a near-litter. We will construct our model in such a way that there is an element x of type α with a $\bot$-extension that is exactly (the atoms inside) N. This will imply, in particular, that each type must have size at least μ by remarks 4.7. Together with the existence of a position function, the size of each type can be seen to be exactly μ. We cannot express the notion of a $\bot$-extension directly with the language available to us, but we do not need it in order to capture the information relevant for this section.

Definition 5.3. Let α be a proper type index with tangle data. We say that α has *typed near-litters* if there is an injection typed_α from the set of near-litters to τ_α, commuting with allowable permutations in the following sense.

$$\rho(\text{typed}_\alpha N) = \text{typed}_\alpha(\rho_{(\alpha > \bot)}(N))$$

5.2 The fuzz map

We will now perform a construction that underpins the rest of the model, and helps us enforce extensionality. In tangled type theory, a given model element has extensions at each level below it. In our model, each object has a 'preferred' extension, and must find a way to compute the other extensions

from that information. In order to do this, we need to be able to convert arbitrary model elements into 'junk' at other levels, which can then be interpreted as a 'non-preferred' extension.

The *fuzz maps* perform this task. They are parametrised by type indices $\beta \neq \gamma$ where $\gamma \neq \bot$, representing the source type level and the target type level. For each such pair, the fuzz map is an injection from β-tangles to litters. An arbitrary litter can only be the image of a fuzz map defined at a single pair of type levels.

Treating the output of the fuzz map as a typed near-litter, its position is always greater than the position of the input to the function. A similar property holds for atoms. This ensures a well-foundedness condition that we can use in many places later.

Definition 5.4. Let β be a type index, and let $\gamma \neq \beta$ be a proper type index. Suppose that β, γ have tangle data and a position function, and that γ has typed near-litters. A *fuzz map* $f_{\beta,\gamma}$ is an injection from τ_β to the set of litters, such that

(i) For any $t \in \tau_\beta$, $f_{\beta,\gamma}(t) = (\nu, \beta, \gamma)$ for some $\nu \in \mu$, so all of the fuzz maps have pairwise disjoint ranges.

(ii) For any $t \in \tau_\beta$ and any near-litter N near $f_{\beta,\gamma}(N)$,

$$\iota_\beta(t) < \iota_\gamma(\text{typed}_\gamma N)$$

(iii) For any $t \in \tau_\beta$ and atom a in the litter set corresponding to $f_{\beta,\gamma}(t)$,

$$\iota_\beta(t) < \iota_\bot(a)$$

We will now show that such a function exists, and will thenceforth refer to it as *the* fuzz map.

Proposition 5.5. Let β be a type index, and let $\gamma \neq \beta$ be a proper type index. Suppose that β, γ have tangle data and a position function, and that γ has typed near-litters. Then a fuzz map $f_{\beta,\gamma}$ exists.

Proof. We construct the map by recursion along τ_β, with the order induced by ι_β. Suppose we have already constructed $f_{\beta,\gamma}$ for $s \in \tau_\beta$ with $\iota(s) < \iota(t)$. We must choose an index $\nu \in \mu$ subject to three constraints:

(i) ν was not picked for any previous s;

(ii) $\iota(t)$ is greater than $\iota_\gamma(\text{typed}_\gamma N)$ for any near-litter near (ν, β, γ);

(iii) $\iota(t)$ is greater than $\iota_\bot(a)$ for any atom a with $a^\circ = (\nu, \beta, \gamma)$.

Each constraint denies us less than μ choices, and so there is always an available index ν to choose. $\square$

5.3 Codes and clouds

From now, we will assume that $\alpha \in \lambda$ is a fixed proper type index at which we intend to construct tangle data. In this section, all other (proper) type indices will be assumed to be strictly less than α. We assume that we have tangle data, position functions, and typed near-litters for all type indices below α.

We will now begin the process of stitching together the fuzz maps to form the function that deduces an alternative extension from a preferred extension.

Definition 5.6. Let $\beta < \alpha$ be a type index, and let $\gamma < \alpha$ be a proper type index not equal to β. The *cloud map* $c_{\beta,\gamma} : \mathcal{P}(\tau_\beta) \to \mathcal{P}(\tau_\gamma)$ is given by

$$c_{\beta,\gamma}(s) = \text{typed}_\gamma{}'' \bigcup_{t \in s} \{N \mid N^\circ = f_{\beta,\gamma}(t)\}$$

For each $t \in s$, the map $c_{\beta,\gamma}$ produces the 'cloud' of near-litters near to $f_{\beta,\gamma}(t)$, and turns them into tangles through the typed near-litter map. This will be used to construct the alternative extension map.

Note. In [4], the cloud map is called $A_{\beta,\gamma}$, but in this paper the notation $A_{\beta,\gamma}$ is reserved for the alternative extension map.

Remark 5.7. The cloud map is injective, and further, if $c_{\beta,\gamma}(s) = c_{\beta',\gamma}(s')$ and s is nonempty, then $\beta = \beta'$. In particular, a nonempty set of γ-tangles has at most one inverse image under any cloud map $c_{\beta,\gamma}$. We will show that there are only finitely many iterated images under an inverse cloud map, and we will use this to define preferred extensions. To prove this, we will make the following definition that encapsulates the notion of a particular choice of $\beta < \alpha$ and an extension at that level.

Definition 5.8. Let $\alpha \in \lambda$. An α-*code* is a pair (β, s) where $\beta < \alpha$ is a type index, and $s \subseteq \tau_\beta$.

Definition 5.9. Let $\beta < \alpha$ be a proper type index. The *code cloud map* C_β is a map from α-codes to α-codes, given by

$$C_\beta((\gamma, s)) = \begin{cases} (\beta, s) & \text{if } \gamma = \beta \\ (\beta, c_{\gamma,\beta}(s)) & \text{if } \gamma \neq \beta \end{cases}$$

That is, C_β applies the relevant cloud map to obtain a code describing a β-extension, or does nothing if the first projection of the code is already β.

Observe that C_β is injective on codes with first projection not equal to β. We will show that there are only finitely many iterated images under the inverse of C_β.

Definition 5.10. Define a relation $\rightsquigarrow$ on α-codes by letting $x \rightsquigarrow C_\beta(x)$ whenever $\text{pr}_1(x) \neq \beta$.

Remark 5.11. A code x has at most one predecessor under $\rightsquigarrow$. If $x \rightsquigarrow y$, then x is empty (more precisely, its second projection is empty) if and only if y is. Moreover, all empty codes are related to each other under $\rightsquigarrow$.

Lemma 5.12. The relation $\rightsquigarrow$ is well-founded on nonempty codes.

Proof. For each nonempty α-code x, we define an ordinal $\iota(x) \in \mu$ given by the smallest position of any tangle in $\text{pr}_2(x)$.

$$\iota((\beta, s)) = \min_{t \in s} \iota_\beta(t)$$

We show that if $x \rightsquigarrow y$, then $\iota(x) < \iota(y)$. Let $x = (\beta, s)$ and $y = (\gamma, c_{\beta,\gamma}(s))$. Suppose $t \in \tau_\gamma$ is the tangle with smallest position in $c_{\beta,\gamma}(s)$. Then $t = \text{typed}_\gamma N$ where $N = f_{\beta,\gamma}(t')$ for some $t' \in s$. It suffices to show $\iota_\beta(t') < \iota_\gamma(t)$, but this holds by the construction of the fuzz map in definition 5.4. □

An object in our model will correspond to a collection of codes, representing its extensions. A given object at level α has exactly one extension at every proper type index $\beta < \alpha$, and may optionally have a $\bot$-extension.

By lemma 5.12, the set of nonempty α-codes forms a tree. Each code has a finite height; a code with no $\rightsquigarrow$-predecessor is of height zero, and the height of any other code can be computed recursively. We

will partition the tree of nonempty α-codes into smaller trees of height 1, and use those as equivalence classes of codes representing a single model element.

Definition 5.13. We define an equivalence relation $\equiv$ on nonempty α-codes, generated by the assertion that $x \equiv y$ whenever $x \rightsquigarrow y$ and x has even height.

Remarks 5.14.   (i) If two codes x, y are equivalent, then either $x = y$, or x has even height and $x \rightsquigarrow y$ (or vice versa), or both x, y have odd height and there is some z with $z \rightsquigarrow x, y$.

  (ii) Each equivalence class contains precisely one even code. This will be the 'preferred' extension of the corresponding model element, as outlined in §3.3.

 (iii) Each equivalence class contains exactly one code for each proper type index $\beta < \alpha$. If the equivalence class contains a code at level $\bot$, it must be the even code, because such codes can never be produced by the code cloud map.

5.4   Model elements

We now collate equivalence classes of codes to create our model elements.

Definition 5.15. An α-*preobject* x assigns to each proper type index $\beta < \alpha$ a set of β-tangles x_β, in such a way that either

  (i) there is a set s of atoms such that $x_\beta = c_{\bot,\beta}(s)$ for all $\beta < \alpha$; or

  (ii) there is a proper type index $\beta < \alpha$ such that (β, x_β) is an even code and $x_\gamma = c_{\beta,\gamma}(x_\beta)$ for all $\gamma \neq \beta$.

Thus nonempty preobjects correspond to equivalence classes of nonempty codes, and there is precisely one empty α-preobject which corresponds to all of the empty codes. An even representative code can be easily extracted from any preobject.

These preobjects satisfy the required form of extensionality in tangled type theory.

Theorem 5.16. Let x, y be α-preobjects, and let $\beta < \alpha$ be a proper type index such that $x_\beta = y_\beta$. Then $x = y$.

Note that applying metatheoretic extensionality to x_β and y_β strengthens this assertion into the proper form.

Proof. First, note that if x and y have the same representative (even) code, then they are equal. This follows from the fact that the extensions of a preobject can be calculated by appling the code cloud map to the representative code. We have three cases.

  (i) *The extensions of x and y can both be calculated from a set of atoms.* Let s_x, s_y be the set of atoms for x, y respectively. Then $c_{\bot,\beta}(s_x) = x_\beta = y_\beta = c_{\bot,\beta}(s_y)$, hence $s_x = s_y$, so x and y have the same representative code.

  (ii) *The extensions of x can be calculated from a set of atoms s, and the extensions of y can be calculated from its γ-extension.* We must show that $(\bot, s) \equiv (\gamma, y_\gamma)$. Suppose $\beta = \gamma$, so $c_{\bot,\beta}(s) = x_\beta = y_\beta$. In this case, $(\bot, s) \equiv (\beta, x_\beta)$ by assumption, giving the result. Instead, suppose $\beta \neq \gamma$. In this case,
$$(\gamma, y_\gamma) \equiv (\beta, c_{\gamma,\beta}(y_\gamma)) = (\beta, y_\beta) = (\beta, x_\beta) \equiv (\bot, s_x)$$
The case where x and y are swapped holds by symmetry.

13

(iii) *The extensions of x can be calculated from its γ-extension, and the extensions of y can be calculated from its δ-extension.* We must show $(\gamma, x_\gamma) \equiv (\delta, y_\delta)$. We have

$$(\gamma, x_\gamma) \equiv (\beta, x_\beta) = (\beta, y_\beta) \equiv (\delta, y_\delta)$$

$\square$

We want to define the α-objects to be those α-preobjects with a small support under the action of α-allowable permutations, so we must first define this group. The allowable permutations will be built up from allowable permutations at lower levels. These will become the derivatives along the one-step paths $\alpha > \beta$. These permutations are constrained so as to commute with the fuzz map.

Definition 5.17. An α-*allowable permutation* ρ assigns to each type index $\beta < \alpha$ a β-allowable permutation $\rho_{(\alpha>\beta)}$, in such a way that

$$(\rho_{(\alpha>\gamma)})_{(\gamma>\perp)}(f_{\beta,\gamma}(t)) = f_{\beta,\gamma}(\rho_{(\alpha>\beta)}(t)) \tag{$*$}$$

for all type indices β, all proper type indices $\gamma \neq \beta$, and all β-tangles t. The homomorphism to the group of α-structural permutations is given by mapping each of the one-step derivatives $\rho_{(\alpha>\beta)}$ to structural permutations and collating the results.

Remark 5.18.   (i) Note that the appearance of β-allowable permutations in this definition is not circular; these are assumed to exist in the tangle data assigned to level β. We will later stipulate that the allowable permutations obtained through tangle data at lower levels already satisfy this condition.

(ii) A small calculation reveals that condition $(*)$ enforces, for any allowable ρ, that for any set $s \subseteq \tau_\beta$ and $\gamma \neq \beta$ proper,

$$\rho_{(\alpha>\gamma)}{}'' c_{\beta,\gamma}(s) = c_{\beta,\gamma}(\rho_{(\alpha>\beta)}{}'' s)$$

Hence, for any code x, we have $\rho(C_\beta(x)) = C_\beta(\rho(x))$, and so allowable permutations preserve code equivalence. This suggests that they are a sensible choice for permutations that respect the structure of α-preobjects.

We are now in a position to define our model elements.

Definition 5.19. An α-*object* is an α-preobject that admits an α-support that supports it under the action of α-allowable permutations. That is, x is an object if there is some support S such that $\rho(S) = S$ implies $\rho(x) = x$. An α-*tangle* is a pair (x, S) where x is an α-object and S is an α-support that supports x under the action of allowable permutations.

We have thus defined tangle data at level α. We can also immediately define typed near-litters at level α, by considering the preobject x given by the even code $(\perp, N)$. This evidently has a small support S given by

$$\operatorname{dom} S = 1; \quad S(0) = ((\alpha > \perp), N)$$

and thus this is an α-object. The map $\operatorname{typed}_\alpha$ thus sends N to the tangle (x, S).

5.5   Comments

We have constructed tangle data and typed near-litters at level α, under the assumption that they exist (along with position functions) at all lower type levels.

The use of a group of permutations is known as a *Fraenkel–Mostowski construction*. Permutations of atoms in a model of a set theory with atoms give rise to permutations of the entire structure. Typically, such constructions involve a normal filter of subgroups of such permutations; this role is played by our allowable permutations.

Recall that a tangle is a model elements bundled together with a support. The main benefit that tangles provide is that the support function supp commutes with allowable permutations. In other Fraenkel–Mostowski constructions, objects typically have a *minimal support*, and these commute with allowable permutations; by bundling supports, we retain this nice behaviour without requiring our objects to have minimal supports (and in fact they do not). Note that tangles themselves do have minimal supports, namely, the support given by supp. Picturing tangles instead of model elements as the primitive notion, we lose extensionality but gain some some symmetry that we will use to great effect later. The idea to bundle tangles in this way was not part of the original proof, and is a new contribution by this author.

We will later impose a restriction on objects that if (x, S) is an element in some extension, then (x, T) is also an element for any other support T. We can then collapse the entire model by taking the first projection of every tangle, reducing it to a model of tangled type theory satisfying extensionality as desired.

The condition that $\mathrm{supp}(\rho(t)) = \rho(\mathrm{supp}\, t)$ was not used in this section. In particular, this entire construction succeeds by replacing the lower-type tangles in the tangle data with model elements themselves. This method was used in an earlier version of the proof before shifting to the new 'bundled' approach.

6 Freedom of action

6.1 Outline

In this section, we will prove the freedom of action theorem. This will allow us to prove that the collection of model elements we have constructed has size exactly μ, among other things. The idea of the theorem is that under suitable hypotheses, a partially defined map of α-addresses can be extended into an α-allowable permutation.

There are many different things one could mean by a 'partially defined map', and we will need several of these for various purposes. The two most important forms of partial maps of addresses are *approximations* and *actions*, the former of which will be used in the statement of the theorem. We will prove some technical results linking approximations and actions.

We will construct the allowable permutation by induction along a particular well-founded relation on addresses: we will define $\rho((A, x))$ given the knowledge of $\rho((B, y))$ for all $(B, y) \prec (A, x)$. In certain cases, we will need to know that the freedom of action theorem holds at all lower levels; we accomplish this by wrapping the entire proof in an induction over all proper type indices below α.

We will then demonstrate that the map as defined is a permutation, and fulfils the criteria to be an allowable permutation.

Our proof for this theorem will not depend on the construction given in §5, but rather on an explicit list of properties that the tangles at each level must satisfy for the theorem to hold; this is discussed in §6.3.

15

6.2 Partial permutations

In the proof of freedom of action, we will often have injective functions that we wish to turn into permutations. In this subsection, we describe a sufficient condition for such a permutation to exist.

Definition 6.1. A *partial permutation* on a set X is a partial function $f : X \dashrightarrow X$ that is a permutation of its domain.

Suppose we have an injective partial function $f : X \dashrightarrow X$. We will construct a pair of functions g and h that agree with f and its inverse respectively on $s = \operatorname{dom} f$, in such a way that forms a partial permutation of X. In particular, consider the diagram

$$\cdots \to L_2 \to L_1 \to L_0 \to s \setminus f''s \overset{f}{\to} \cdots \overset{f}{\to} f''s \setminus s \to R_0 \to R_1 \to R_2 \to \cdots$$

To fill in the orbits of f, we construct a sequence of disjoint subsets of X called L_n and R_n, where for each $i : \mathbb{N}$, the cardinality of L_i is the same as $s \setminus f''s$, and the cardinality of R_i is that of $f''s \setminus s$. There are natural bijections along this diagram, mapping L_{n+1} to L_n and R_n to R_{n+1}. There are also bijections $f''s \setminus s \to R_0$ and $L_0 \to s \setminus f''s$. This completes the diagram, giving a partial permutation extending f defined on

$$s \cup f''s \cup \left(\bigcup_{i \in \mathbb{N}} L_i \right) \cup \left(\bigcup_{i \in \mathbb{N}} R_i \right)$$

6.3 Hypotheses

Let α be a proper type index. The complete list of hypotheses we will use when proving the freedom of action theorem at all levels $\beta \le \alpha$ are as follows.

 (i) We assume that we have tangle data and typed near-litters at all proper type levels $\beta \le \alpha$. The data for level α will later be provided by running the construction in §5. We also assume position functions at all proper type levels $\beta < \alpha$.

 (ii) We assume that the groups of allowable permutations A_β can be related by the use of a derivative map. Each path $A : \beta \rightsquigarrow \gamma$ gives rise to a group homomorphism $A_\beta \to A_\gamma$, and its action on an allowable ρ is denoted ρ_A. We assume that the following diagram commutes.

$$
\begin{array}{ccc}
A_\beta & \longrightarrow & S_\beta \\
{\scriptstyle \rho \mapsto \rho_A} \downarrow & & \downarrow {\scriptstyle \pi \mapsto \pi_A} \\
A_\gamma & \longrightarrow & S_\gamma
\end{array}
$$

Alternatively, viewing allowable permutations as subgroups of structural permutations, this assumption states that any derivative of an allowable permutation is also allowable.

 (iii) Let $\beta < \alpha$ be a proper type index and let $t \in \tau_\beta$. Let $\gamma < \alpha$ be a type index, let $\delta < \alpha$ be a proper type index distinct from γ, and let $s \in \tau_\gamma$. Suppose that $(A, x) \in \operatorname{ran} \operatorname{supp} t$, where A is a β-extended type index and x is an atom or a near-litter. Suppose further that either $x = a$ is an atom with $a^\circ = f_{\gamma,\delta}(s)$, or that $x = N$ is a near-litter that has nonempty intersection with $\mathsf{LS}(f_{\gamma,\delta}(s))$. Then $\iota_\gamma(s) < \iota_\beta(t)$.

 (iv) Let $\beta \le \alpha$ be a type index, let $\gamma < \alpha$ be a type index, and let $\delta < \alpha$ be a proper type index distinct from γ. Let ρ be a β-allowable permutation, and let t be a γ-tangle. Then

$$\rho_{(\beta > \delta > \bot)}(f_{\gamma,\delta}(t)) = f_{\gamma,\delta}(\rho_{(\beta > \gamma)}(t))$$

This is a restatement of equation $(\ast)$ from definition 5.17.

(v) Conversely, let $\beta \leq \alpha$ be a type index, and let $\rho_{(\beta > \gamma)}$ be a γ-allowable permutation for each type index $\gamma < \beta$. Suppose that the $\rho_{(\beta > \gamma)}$ satisfy the one-step fuzz constraints

$$(\rho_{(\gamma > \delta)})_{(\delta > \perp)}(f_{\gamma, \delta}(t)) = f_{\gamma, \delta}(\rho_{\beta > \gamma}(t))$$

Then we can assemble the $\rho_{(\beta > \gamma)}$ into a β-allowable permutation; more precisely, there exists an allowable permutation ρ with one-step derivatives given by the $\rho_{(\beta > \gamma)}$.

Remarks 6.2.   (i) Apart from (iii), these conditions are obviously satisfied by our construction in §5. Note that this remaining condition only constrains type indices strictly below α, so we may carry this knowledge through the main inductive step. Once we know that a given type is of size μ, a position function can be constructed in such a way to make (iii) hold, using a similar method to the construction of the fuzz map in definition 5.4.

(ii) Hypothesis (v) will be used exactly once, at the end of freedom of action. This will allow us to build the allowable permutation asserted in the theorem.

6.4   Dependency of addresses

As described in §6.1, we will define a well-founded relation on the set of α-addresses that will allow us to define the action of an allowable permutation recursively.

Definition 6.3. Let $\beta \leq \alpha$ be a proper type index. We define the *dependency* relation $\prec$ on β-support conditions by the following constructors.

(i) Let A be a β-extended index, and let a be an atom. Then

$$(A, \mathsf{NL}(a^\circ)) \prec (A, a)$$

(ii) Let A be a β-extended index, and let N be a near-litter with $N \neq \mathsf{NL}(N^\circ)$. Then

$$(A, \mathsf{NL}(N^\circ)) \prec (A, N)$$

(iii) Let A be a β-extended index, let N be a near-litter, and let a be an atom in the symmetric difference $N \,\triangle\, \mathsf{LS}(N^\circ)$. Then

$$(A, a) \prec (A, N)$$

(iv) Let $\gamma \leq \alpha$ be a proper type index, and let $\delta, \varepsilon < \gamma$ be distinct proper type indices. Let A be a path $\beta \rightsquigarrow \gamma$ and let $t \in \tau_\delta$. Then for all $c \in \mathrm{ran}\,\mathrm{supp}\,t$,

$$(A \gg (\gamma > \delta) \gg \mathrm{pr}_1(c), \mathrm{pr}_2(c)) \prec (A \gg (\gamma > \varepsilon > \perp), \mathsf{NL}(f_{\delta, \varepsilon}(t)))$$

(v) Let $\gamma \leq \alpha$ be a proper type index, and let $\varepsilon < \gamma$ be a proper type index. Let A be a path $\beta \rightsquigarrow \gamma$ and let a be an atom. Then

$$(A \gg (\gamma > \perp), a) \prec (A \gg (\gamma > \varepsilon > \perp), \mathsf{NL}(f_{\perp, \varepsilon}(a)))$$

If $x \prec y$, we say that y *depends on* x.

17

Remarks 6.4.　　(i) Let x be an address and ρ be an allowable permutation we wish to study, and suppose we know the values of $\rho(y)$ for $y \prec x$. We can then deduce some useful information about the possible values of $\rho(x)$, and in some cases determine it exactly (for instance, if x is a near-litter $N \neq \mathsf{NL}(N°)$).

 (ii) Each address depends on only a small set of other addresses. This holds as supports have small ranges.

(iii) A litter is said to be *inflexible* with respect to a particular β-extended index if it is the image of a fuzz map as given by constructors (iv) or (v) above. A litter is *flexible* with respect to an extended index if it is not inflexible. The addresses formed from flexible litters are minimal elements under $\prec$. It is technically possible for an address formed from an inflexible litter to be minimal under $\prec$ if its support is empty, but aside from this pathology, this is a full description of the minimal elements.

Proposition 6.5. The dependency relation is well-founded.

Proof. If X is a set endowed with a relation $\prec$, we say that $x \in X$ is $\prec$-*accessible* if all of its $\prec$-predecessors are $\prec$-accessible. A relation on X is well-founded precisely when every $x \in X$ is accessible under that relation. We make the following observations about accessibility of addresses under the dependency relation, both of which follow directly from the relevant definitions.

 (i) Let A be a β-extended type index, and let a be an atom. Then (A, a) is accessible if $(A, \mathsf{NL}(a°))$ is accessible.

 (ii) Let A be a β-extended type index, and let N be an near-litter. Then (A, N) is accessible if $(A, \mathsf{NL}(a°))$ is accessible for all $a \in N$.

We have thus reduced accessibility of arbitrary addresses to accessibility of litter addresses. Motivated by this, we will define an auxiliary relation R on litters and prove that it is well-founded. We define the following constructors for R.

 (i) Let $\delta, \varepsilon < \alpha$ be distinct proper type indices, and let $t \in \tau_\delta$. Then if B is an extended index and a is an atom such that either $(B, a) \in \operatorname{ran} \operatorname{supp} t$ or $(B, N) \in \operatorname{ran} \operatorname{supp} t$ and $a \in N$, then

$$a° \, R \, f_{\delta,\varepsilon}(t)$$

 (ii) Let $\varepsilon < \alpha$ be a proper type index, and let a be an atom. Then

$$a° \, R \, f_{\perp,\varepsilon}(a)$$

We will say that a litter L has position $\nu \in \mu$ if either

 (i) $\delta, \varepsilon < \alpha$ are distinct proper type indices, $t \in \tau_\delta$, and

$$L = f_{\delta,\varepsilon}(t); \quad \nu = \iota(t)$$

　　or

 (ii) $\varepsilon < \alpha$, a is an atom, and

$$L = f_{\perp,\varepsilon}(a); \quad \nu = \iota(a)$$

18

First, note that a litter has at most one position; this follows by injectivity of the fuzz map. It follows by definition that if $L_1 \mathrel{R} L_2$, then L_2 has a position. Moreover, if additionally L_1 has a position, its position is less than that of L_2. This last fact follows from part (iii) of the definition of the fuzz map in definition 5.4, as well as hypothesis (iii) in §6.3. In particular, as the order on μ is well-founded, the relation R must also be well-founded.

We show by induction on R that for every litter L and β-extended index A, the address $(A, \mathsf{NL}(L))$ is accessible; the result then follows from the observations above, since they concern only litter addresses. Suppose $c \prec (A, \mathsf{NL}(L))$; we must show c is accessible. Only cases (iv) and (v) of definition 6.3 can possibly occur. Suppose $c = (B, a)$ and $L = f_{\delta,\varepsilon}(t)$. Then by observation (i), it suffices to show $(B, \mathsf{NL}(a^{\circ}))$ is accessible, but this is given by the inductive hypothesis. If $c = (B, N)$ and $L = f_{\delta,\varepsilon}(t)$, then the result similarly holds by applying observation (ii) to the inductive hypothesis. Finally, if $c = (B, a)$ and $L = f_{\perp,\varepsilon}(a)$, then again the result holds by observation (i) and the inductive hypothesis. $\qquad\square$

Remark 6.6. If $x \prec y$ are γ-addresses and B is a path $\beta \rightsquigarrow \gamma$, then

$$(B \gg \mathsf{pr}_1(x), \mathsf{pr}_2(x)) \prec (B \gg \mathsf{pr}_1(y), \mathsf{pr}_2(y))$$

We will define a partial order on addresses by defining $<$ to be the transitive closure of $\prec$. Evidently, the relation $<$ is also well-founded.

6.5 Approximations and the theorem statement

We can now define one notion of a partially defined map of addresses.

Definition 6.7. A *near-litter approximation* is a pair $\varphi = (\varphi^A, \varphi^L)$ where φ^A is a partial permutation of atoms, φ^L is a partial permutation of litters, and the domain of φ^A contains only a small number of atoms associated to a given litter L.

For brevity, we will often suppress the superscripts on approximations.

Definition 6.8. A near-litter approximation φ is said to *approximate* a near-litter permutation π when they agree at all atoms and litters at which φ is defined.

$$\forall a \in \operatorname{dom} \varphi^A.\, \pi(a) = \varphi(a); \quad \forall L \in \operatorname{dom} \varphi^L.\, \pi(L) = \varphi(L)$$

An atom a is said to be *exceptional* in a near-litter permutation π if

$$(\pi(a))^{\circ} \neq \pi(a^{\circ}) \quad \text{or} \quad (\pi^{-1}(a))^{\circ} \neq \pi^{-1}(a^{\circ})$$

We say that φ *exactly* approximates π if it approximates π, and in addition, all exceptional atoms in π are in $\operatorname{dom} \varphi^A$.

Remark 6.9. If φ exactly approximates π, we can determine the exact image of $\pi(\mathsf{NL}(L))$ for any litter $L \in \operatorname{dom} \varphi^L$ from data in φ. It can be shown directly that every near-litter approximation exactly approximates some near-litter permutation, although we will not need this fact.

In the same way that structural permutations were built up from near-litter permutations in definition 4.12, we will build up structural approximations from near-litter approximations.

Definition 6.10. Let β be a type index. A β-*structural approximation* φ assigns a near-litter approximation to each β-extended index. For a given β-extended index A, the assigned near-litter permutation is denoted φ_A. A β-structural approximation φ is said to *(exactly) approximate* a β-structural

permutation π if φ_A (exactly) approximates π_A for all β-extended indices A. We say that β is *free* if for all β-extended indices A, the only litters that appear in $\text{dom}(\varphi_A)^L$ are A-flexible.

We can now state the freedom of action theorem.

Theorem 6.11 (freedom of action)**.** Let $\beta \leq \alpha$ be a proper type index. Then any free β-structural approximation φ exactly approximates some β-allowable permutation ρ.

Remarks 6.12.   (i) This theorem extends the fact that any near-litter approximation exactly approximates some near-litter permutation; in fact, this statement is precisely freedom of action for $\beta = \bot$.

  (ii) If L is A-inflexible, then the value of $\rho_A(L)$ depends on a the value of a certain tangle under ρ, so we may not choose it freely. The freedom of action theorem states that this is the only restriction to the images of litters under allowable permutations.

 (iii) The theorem is phrased in terms of approximations, although this is somewhat arbitrary. The purpose of the theorem is to construct allowable permutations, but the data we need to encapsulate in an allowable permutation may come in various different forms.

 (iv) As described in §3.4, we will later use the freedom of action theorem to show that the set of α-objects has cardinality exactly μ.

6.6   Actions

We will now introduce another form of data that can be used as the input to the freedom of action theorem.

Definition 6.13. A *near-litter action* is a pair $\psi = (\psi^A, \psi^L)$ where ψ^A is a partial function from the set of atoms to itself, ψ^L is a partial function from the set of litters to the set of near-litters, and both ψ^A and ψ^L have small domain.

As with approximations, we will suppress the superscripts when possible.

Remarks 6.14.   (i) The function ψ^L encapsulates the mapping $L \mapsto \pi(\text{NL}(L))$, whereas for near-litter approximations φ, the function φ^L encapsulates $L \mapsto \pi(L)$. In this way, near-litter actions keep track of more data about where a litter is sent under a permutation, since we know precisely where its set of atoms is sent.

  (ii) The appropriate notion of approximating a permutation here is that

$$\forall a \in \text{dom}\,\psi^A.\,\pi(a) = \psi(a); \quad \forall L \in \text{dom}\,\psi^L.\,\pi(\text{NL}(L)) = \psi(L)$$

We need no distinction between approximating and exactly approximating since near-litter actions already capture precise images of litters. Not every near-litter action approximate some near-litter permutation, because we may have a situation where $a \in \text{dom}\,\psi^A, a^\circ \in \text{dom}\,\psi^L$ but $\psi(a) \notin \psi(a^\circ)$. This pathology, as well as injectivity, turns out to be the only way in which near-litter actions can fail to approximate some near-litter permutation. We therefore make the following definition.

Definition 6.15. A near-litter action ψ is said to be *lawful* if

  (i) if a_1, a_2 are atoms and $\psi(a_1) = \psi(a_2)$ then $a_1 = a_2$;

  (ii) if L_1, L_2 are litters and $\psi(L_1) \cap \psi(L_2) \neq \varnothing$ then $L_1 = L_2$;

 (iii) if $a \in \text{dom}\,\psi^A$ and $L \in \text{dom}\,\psi^L$, then $a^\circ = L$ if and only if $\psi(a) \in \psi(L)$.

20

We now describe a procedure for converting lawful near-litter actions into free near-litter approximations. This process has two steps: expanding the near-litter action so that it is *precise*, and then completing it into an approximation.

Definition 6.16. A near-litter action ψ is said to be *precise* if for all $L \in \operatorname{dom} \psi^L$,

(i) $\psi(L) \triangle \mathsf{LS}(\psi(L)^\circ) \subseteq \operatorname{ran} \psi^A$;

(ii) $\operatorname{ran} \psi^A \cap \mathsf{LS}(L) \subseteq \operatorname{dom} \psi^A$;

(iii) $\operatorname{dom} \psi^A \cap \psi(L) \subseteq \operatorname{ran} \psi^A$.

These are precisely the requirements that our completion process needs in order to compute complete orbits for all atoms and litters.

Lemma 6.17. Any lawful near-litter action can be extended to a precise near-litter action with the same map on litters.

This proof is explicit, but unenlightening.

Proof. Let ψ be a near-litter action. Call a litter *banned* for ψ if it contains any atom in the domain or range of ψ^A, is in $\operatorname{dom} \psi^L$, or shares an atom with any near-litter in $\operatorname{ran} \psi^L$. Observe that there are only a small amount of banned litters for a given ψ. We will use the unbanned litters to store the extra atoms we need to insert into ψ.

We will first extend ψ so that it satisfies condition (i) in definition 6.16. Suppose that $a \in \mathsf{LS}(\psi(L)^\circ) \setminus \psi(L)$; we need to allocate a preimage for a under ψ. Choose an atom a' in an unbanned litter, and add to ψ^A that $\psi(a') = a$. The addition of a' to ψ does not create more problems with satisfying definition 6.16, since a' is part of an unbanned litter. There are only a small amount of atoms in such circumstances, so they can all be added to the same unbanned litter for convenience.

Now suppose that $a \in \psi(L) \setminus \mathsf{LS}(\psi(L)^\circ)$. We need to allocate a preimage for a under ψ that lies in L. Since there are only a small amount of atoms in question, and only a small amount of atoms in L already appear in $\operatorname{dom} \psi^A$, we can fit all necessary preimages inside L. Again, this does not create any more problematic atoms that we must deal with. Performing this process for every $L \in \operatorname{dom} \psi^L$, we obtain a new near-litter action ψ' which satisfies condition (i) of definition 6.16 and has the same litter map as ψ. One can easily check that ψ' is lawful.

We now extend ψ' so that it satisfies conditions (ii) and (iii), while not disturbing condition (i). To do this, we need to fill in certain images and preimages of atoms. First, use the construction in §6.2 to produce a partial permutation of litters π extending $L \mapsto \psi'(L)^\circ$. We require that all extra litters used to make $\operatorname{dom} \pi$ are unbanned; this can be arranged since we have a choice of μ unbanned litters to use. We then extend π to be defined on all banned litters by setting $\pi(L) = L$ in this case; the resulting map is clearly still a partial permutation. This map will be used to control the litters inside which the new images and preimages of atoms under ψ' will be allocated. For now, we will describe only the allocation of forward images; backward images work entirely symmetrically.

Inside each litter, we reserve enough unused atoms to correspond with all possible pairs (n, a) where $a \in \operatorname{ran} \psi^A \setminus \operatorname{dom} \psi^A$. We then define that an atom $a \in \operatorname{ran} \psi^A \setminus \operatorname{dom} \psi^A$ is mapped to the atom corresponding to $(0, a)$ in the litter $\pi(a^\circ)$. If a' is an atom corresponding to (n, a) in the litter $\pi^{n+1}(a^\circ)$, then it is mapped to the atom corresponding to $(n + 1, a)$ in the litter $\pi^{n+2}(a^\circ)$. This behaviour ensures that we our extra atoms do not violate condition (i) of definition 6.16. Note that most atoms corresponding to one of the pairs (n, a) will not have an image defined. This completes the description

21

of the forward orbits; after filling the backward orbits analogously, the resulting near-litter action ψ'' can be seen to be precise and lawful. □

We now describe the method for turning a precise action into a free approximation.

Lemma 6.18. Let ψ be a precise lawful near-litter action and let A be a β-extended index. Suppose that ψ maps A-flexible litters to near-litters that are near A-flexible litters. Then there is a near-litter approximation φ whose litter map contains only A-flexible litters, such that if φ exactly approximates π, we have

(i) if $a \in \operatorname{dom} \psi^A$, then $\psi(a) = \pi(a)$;

(ii) if $L \in \operatorname{dom} \psi^L$ is A-flexible, then $\psi(L)° = \pi(L)$; and

(iii) if $L \in \operatorname{dom} \psi^L$ and $\psi(L)° = \pi(L)$, then $\psi(L) = \pi(\mathsf{NL}(L))$.

This lemma provides exactly what is needed to use the freedom of action theorem with actions instead of approximations directly. We will shortly restate this lemma in a more useful form.

Proof. First, consider the restriction of ψ^L to the A-flexible litters. This map can be completed to a partial permutation φ^L as in §6.2, where any added litters are flexible and unbanned; there is a supply of μ such litters, so this can always be arranged. We similarly complete ψ^A into a partial permutation φ^A; the added atoms are chosen to belong to another unbanned litter. This yields a near-litter approximation $\varphi = (\varphi^A, \varphi^L)$. We claim that this satisfies the conclusion of the lemma. It is clear that (i) and (ii) hold. Part (iii) can be shown by checking all possible cases of where an atom can come from and be sent; this requires the full strength of the hypothesis that ψ was precise. □

This concludes the main work of this subsection. We now briefly describe how near-litter actions are composed to form structural actions, and combine lemmas 6.17 and 6.18 to make an easy way to apply freedom of action to structural actions.

Definition 6.19. Let β be a type index. A β-*structural action* ψ assigns a near-litter action to each β-extended index. For a given β-extended index A, the assigned near-litter action is denoted ψ_A. We say that β is *lawful* if for all β-extended indices A, the near-litter action ψ_A is lawful.

Proposition 6.20 (converting actions to approximations)**.** Let ψ be a lawful β-structural action. Suppose that for every β-extended index A, the near-litter action ψ_A maps A-flexible litters to near-litters that are near A-flexible litters. Then there is a free β-structural approximation such that if φ exactly approximates π, we have

(i) if $a \in \operatorname{dom}(\psi_A)^A$, then $\psi_A(a) = \pi_A(a)$;

(ii) if $L \in \operatorname{dom}(\psi_A)^L$ is A-flexible, then $\psi_A(L)° = \pi_A(L)$; and

(iii) if $L \in \operatorname{dom}(\psi_A)^L$ and $\psi_A(L)° = \pi_A(L)$, then $\psi(L) = \pi_A(\mathsf{NL}(L))$.

Proof. Follows directly from lemmas 6.17 and 6.18. □

6.7 Building the permutation

We now begin to prove the freedom of action theorem. Fix a proper type index $\beta \leq \alpha$ and a β-structural approximation φ. We may assume by induction on λ that the freedom of action holds for all proper type indices $\gamma < \beta$; we will use this in §6.7.3.

We will construct the functions $\rho_A^\star$ which map atoms to atoms, litters to litters, and near-litters to near-litters, with the intent that they will be the derivatives of an allowable permutation ρ. We will proceed by recursion along the dependency relation from definition 6.3; at each stage (A, x), we assume that we know the value of $\rho_B^\star(y)$ for all $(B, y) \prec (A, x)$, and compute $\rho_A^\star(x)$. This construction will take a different form depending on what kind of address (A, x) is. In particular, if $x = \mathrm{NL}(L)$ is a near-litter that came from a litter, we will break the derivation into two stages: we will first compute $\rho_A^\star(L)$, and then calculate $\rho_A^\star(\mathrm{NL}(L))$ separately.

6.7.1 Atoms

Suppose we are at stage (A, a) where a is an atom. If a was already assigned by φ, we simply define

$$\rho_A^\star(a) = \varphi_A(a)$$

Suppose $a \notin \mathrm{dom}(\varphi_A)^A$. For each extended index A and litter L, the set $L^- = L \setminus \mathrm{dom}(\varphi_A)^A$ is a near-litter, near the litter L. For each pair of litters L_1, L_2, we choose in advance a bijection π_{L_1, L_2} between L_1^- and L_2^-. We use this bijection now to define

$$\rho_A^\star(a) = \pi_{a^\circ, \rho_A^\star(a^\circ)}(a)$$

6.7.2 Flexible litters

Suppose we are at stage $(A, \mathrm{NL}(L))$ where L is A-flexible. If $L \in \mathrm{dom}(\varphi_A)^L$, then we must define

$$\rho_A^\star(L) = \varphi_A(L)^\circ$$

Otherwise, we will define that

$$\rho_A^\star(L) = L$$

6.7.3 Inflexible litters from proper levels

Suppose we are at stage

$$x = (A \gg (\gamma > \varepsilon > \bot), \mathrm{NL}(f_{\delta, \varepsilon}(t)))$$

where $\delta, \varepsilon < \gamma$ are distinct proper type indices, A is a path $\beta \rightsquigarrow \gamma$, and $t \in \tau_\delta$. We need to enforce the coherence condition that

$$\rho_{(A \gg (\gamma > \varepsilon > \bot))}(f_{\delta, \varepsilon}(t)) = f_{\delta, \varepsilon}(\rho_{(A \gg (\gamma > \delta))}(t))$$

Thus, we need to compute the right-hand side of this equation. However, we do not know if $\rho^\star$ happens to come from an allowable permutation, so there is no way to apply it to an arbitrary tangle t. To deal with this, we invoke the freedom of action theorem at level δ to produce a new allowable permutation ρ' with the same action on t as $\rho_{(A \gg (\gamma > \delta))}$ will eventually be shown to have.

The values of $\rho^\star$ at all addresses that x depends on can be assembled into a β-structural action ψ. We will consider its derivative $\psi_{(A \gg (\gamma > \delta))}$, which is a δ-strutural action. Outside of this recursion, we will be able to show that $\psi_{(A \gg (\gamma > \delta))}$ is lawful and maps flexible litters to flexible litters as required for proposition 6.20; for now, if it turned out not to be lawful, we would simply assign

$$\rho_{A \gg (\gamma > \varepsilon > \bot)}^\star(f_{\delta, \varepsilon}(t)) = f_{\delta, \varepsilon}(t)$$

but this case will never actually occur. Build a δ-structural approximation from $\psi_{(A \gg (\gamma > \delta))}$ as in proposition 6.20, and apply the freedom of action theorem at level $\delta < \gamma \leq \beta$ to obtain a δ-allowable permutation ρ'. We will later be able to show that ρ' agrees with $\rho^\star$ on all elements of the support of t. We set

$$\rho_{A \gg (\gamma > \varepsilon > \bot)}^\star(f_{\delta, \varepsilon}(t)) = f_{\delta, \varepsilon}(\rho'(t))$$

23

6.7.4 Inflexible litters from the base level

Suppose we are at stage

$$x = (A \gg (\gamma > \varepsilon > \bot), \mathsf{NL}(f_{\bot,\varepsilon}(a)))$$

where $\varepsilon < \gamma$ is a proper type index, A is a path $\beta \rightsquigarrow \gamma$, and a is an atom. We must enforce

$$\rho_{(A \gg (\gamma > \varepsilon > \bot))}(f_{\bot,\varepsilon}(a)) = f_{\bot,\varepsilon}(\rho_{(A \gg (\gamma > \bot))}(a))$$

so we have no choice but to define

$$\rho^\star_{A \gg (\gamma > \varepsilon > \bot)}(f_{\bot,\varepsilon}(a)) = f_{\bot,\varepsilon}(\rho^\star_{(A \gg (\gamma > \bot))}(a))$$

6.7.5 Near-litters

Suppose we are at stage (A, N) where N is a near-litter. Either by inductive hypothesis or by sections 6.7.2 to 6.7.4, we know the value of $\rho^\star_A(N^\circ)$. We then define

$$\mathsf{pr}_2(\rho^\star_A(N)) = \left(\rho^\star_A(N^\circ)^- \cup {\psi_A}'' \left(N \cap \mathsf{dom}(\psi_A)^A \right) \right) \bigtriangleup \left\{ \rho^\star_A(a) \mid a \in (N \bigtriangleup \mathsf{LS}(N^\circ)) \setminus \mathsf{dom}(\psi_A)^A \right\}$$

In the case that $N = \mathsf{NL}(L)$, this simplifies to

$$\mathsf{pr}_2(\rho^\star_A(\mathsf{NL}(L))) = \rho^\star_A(L)^- \cup {\psi_A}'' \left(\mathsf{LS}(L) \cap \mathsf{dom}(\psi_A)^A \right)$$

This definition has the property that

$$\rho^\star_A(N) = \rho^\star_A(\mathsf{NL}(N^\circ)) \bigtriangleup {\rho^\star_A}''(N \bigtriangleup \mathsf{LS}(N^\circ))$$

and that for any two distinct litters L_1, L_2, the images $\rho^\star_A(\mathsf{NL}(L_1))$ and $\rho^\star_A(\mathsf{NL}(L_2))$ are disjoint.

6.8 Injectivity

We have fully described $\rho^\star$; we will now show that the $\rho^\star_A$ are near-litter permutations. The first step in this process is proving injectivity of the $\rho^\star_A$.

References

[1] Mario Carneiro. *The Type Theory of Lean*. 2019. URL: https://github.com/digama0/lean-type-theory/releases.

[2] The mathlib Community. "The Lean Mathematical Library". In: *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs*. CPP 2020. New Orleans, LA, USA: Association for Computing Machinery, 2020, pp. 367–381. ISBN: 9781450370974. DOI: 10.1145/3372885.3373824. URL: https://github.com/leanprover-community/mathlib4.

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

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

[5] 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.

[6]   Leonardo de Moura and Sebastian Ullrich. "The Lean 4 Theorem Prover and Programming Language". In: *Automated Deduction – CADE 28: 28th International Conference on Automated Deduction, Virtual Event, July 12–15, 2021, Proceedings.* Berlin, Heidelberg: Springer-Verlag, 2021, pp. 625–635. ISBN: 978-3-030-79875-8. DOI: 10.1007/978-3-030-79876-5_37. URL: https://doi.org/10.1007/978-3-030-79876-5_37.

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

[8]   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.

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

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