Universes and Univalence in Homotopy Type Theory

James Ladyman and Stuart Presnell
University of Bristol

Friday 18th November, 2016

Abstract

The Univalence axiom, due to Vladimir Voevodsky, is often taken to be one of the most important new discoveries arising from the Homotopy Type Theory (HoTT) research programme. It is said by Steve Awodey that Univalence embodies mathematical structuralism, and that Univalence may be regarded as 'expanding the notion of identity to that of equivalence'. In this paper we explore the conceptual, foundational and philosophical status of Univalence in Homotopy Type Theory. We extend our Types-as-Concepts interpretation of HoTT to Universes, and offer an account of the Univalence axiom in such terms. We consider Awodey's informal argument that Univalence is motivated by the principle that reasoning should be invariant under isomorphism, and we examine whether an autonomous and rigorous justification along these lines can be given. We consider two problems facing such a justification. First, there is a difference between *equivalence* and *isomorphism* and Univalence must be formulated in terms of the former. Second, the argument as presented cannot establish Univalence itself but only a weaker version of it, and must be supplemented by an additional principle. The paper argues that the prospects for an autonomous justification are promising.

Contents

1 Introduction

In mathematical practice, if two structures have been shown to be isomorphic they will often be treated as a single entity under two different guises. However, in standard foundational systems arbitrary isomorphic entities are not identical, and formally the notions of identity and isomorphism are distinguished. One of the main innovations of Homotopy Type Theory, a new proposed foundation for mathematics, is the Univalence axiom due to Vladimir Voevodsky [Voevodsky, 2009; Awodey et al., 2013].[1] In a Univalent theory, *identity* and *isomorphism* (or rather, a related notion called 'equivalence'; see Section 6) are regarded as equivalent notions in a precise sense to be explained below. For example in such a theory the Klein four-group and the product $\mathbb{Z}_2 \times \mathbb{Z}_2$ of two copies of the cyclic group of order 2 are not just isomorphic but are identical.

Univalence is taken to be one of the most important new discoveries arising from the 'Homotopy Type Theory/Univalent Foundations' research program initiated by Vladimir Voevodsky and Steve Awodey. It is said by Steve Awodey that Univalence embodies mathematical structuralism: since isomorphism is 'sameness of structure', if Univalence says that isomorphic entities are identical then it entails "that two mathematical objects are identical if and only if they have the same structure [...] In other words, mathematical objects simply *are* structures." [Awodey, 2014b, pp. 10-11]

This should not be thought of as 'collapsing isomorphism to identity', as that would lead to a theory in which there are no non-trivial isomorphisms. Moreover, an important lesson from category theory is that collapsing isomorphic structures by identifying them tends to throw away important information – hence Baez & Dolan's injunction "*never mistake equivalence for equality*" [Baez and Dolan, 1998] (which on the face of it appears to be in direct opposition to Univalence). Rather, the intended interpretation of Univalence is, according to Awodey, to regard it as "*expanding* the notion of identity to that of equivalence." [Awodey, 2014b, p. 9]

In this paper we explore the conceptual, foundational and philosophical status of the Univalence axiom in Homotopy Type Theory. In particular, we address the question of whether Univalent HoTT (that is, HoTT with the Univalence axiom) can be given an autonomous presentation as a foundation for mathematics in the sense defined in [Ladyman and Presnell, 2016].

[1] We will assume familiarity with HoTT, as set out in [The Univalent Foundations Program, 2013] (henceforth the 'HoTT Book'). For a more basic introduction to the language of HoTT, see [Ladyman and Presnell, 2014].

1.1 The justification of Univalence

A mathematician might say that Univalence, like any other axiom, needs no justification beyond its interest or its usefulness. However, if HoTT is to be a foundation for the whole of mathematics (in the sense explained in [Ladyman and Presnell, 2016], in which a 'foundation' goes beyond merely a language or framework in which mathematics can be developed) then each of its definitions, rules, and axioms must be explained and justified in a way that does not require appeal to concepts that must be spelled out using pre-existing mathematics, or to connections with other sophisticated branches of mathematics, or to the intuitions of mathematicians. Moreover, for these purposes the motivation offered for a given axiom must justify that choice of axiom in particular, and not something weaker that follows as a consequence of the axiom. The rules of basic HoTT can all be explained and motivated from pre-mathematical principles (as demonstrated in [Ladyman and Presnell, 2016]), which supports the claim that basic HoTT can serve as an autonomous foundation for mathematics. What remains, then, is to provide a justification for Univalence.

One approach to justifying Univalence begins with the observation that in HoTT identity is always *identity of tokens in a type*, which may be understood as identity *qua* some structure. For example, the Euclidean plane and the Hyperbolic disc are not identical *qua* metric space, the corresponding tokens of the type topological space are identical in that type. The things we can say within the language of HoTT about some entity depend upon what type that entity belongs to, i.e. upon what kind of structure it is being regarded as. Hence one might think that Univalence could be justified as an expression of the Principle of the Identity of Indiscernibles: briefly, that facts about the identity of tokens of a type are exhausted by, and may be reduced to, facts about what can be predicated of those tokens. [Ladyman and Presnell, 2015a] considers the relationship between identity and indiscernibility in HoTT and argues that a re-interpretation of identity types as denoting indiscernibility cannot be sustained, and thus cannot be the basis for a justification of Univalence.

Another approach is to change the formal system itself. There are approaches within the broader HoTT research programme (for example, the 'Cubical Type Theory' of [Bezem et al., 2014]) that define a different but related type theory to the one considered here, and in which Univalence may be proved as a theorem rather than posited as an axiom. One could, then, alternatively seek to give an autonomous presentation of such a system. This is beyond the scope of the present discussion.

In this paper we consider the informal argument given by Awodey in [2014b] (which, to be clear, is not intended to offer a justification of the above kind; see Section 7.2 for further discussion) that Univalence is motivated by the "Principle of Structuralism" that "isomorphic objects are iden-

tical", and thus that reasoning should be invariant under isomorphism.[2] We examine whether an autonomous argument for Univalence can be produced by developing a more rigorous version of Awodey's argument, and consider two problems that must be overcome in doing so.

1.2 Outline of the paper

Since the Univalence axiom is formulated in terms of universes, the next section explicates the notion of universes making explicit the conceptual features they are taken to have in [HoTT Book, Section 1.3]. It also explains how to understand universes in terms of our Types-as-Concepts interpretation [Ladyman and Presnell, 2016]. The formal definition of Univalence is given in Section 3, with one detail suppressed. Section 4 briefly reviews some of the most important formal consequences of Univalence.

Following the account of the criteria for a foundation for mathematics outlined in [Ladyman and Presnell, 2016] (involving five interrelated components: a framework, semantics, metaphysics, epistemology, and methodology) Section 5 considers the meaning of Univalence in relation to the explanation of universes given in Section 2. In Section 5.1, consideration of mathematical metaphysics leads to a possible motivation for adopting Univalence which, on closer inspection, turns out to be inadequate for the purposes of this paper.

Section 6 examines how to fill in the detail that was suppressed in Section 3, namely the definition of 'equivalence' involved in the statement of Univalence, and Section 6.5 observes that if Univalence is formulated in terms of isomorphism (as commonly understood) this leads to inconsistency. This adds a complication to any attempt to give an autonomous justification for Univalence.

Section 7 examines Awodey's informal motivation of Univalence in terms of an Invariance Principle [2014b], and discusses the obstacles that must be overcome in order to produce from it a rigorous justification for the Univalence axiom that could form part of an autonomous foundation for mathematics.

Section 8 suggests that rather than attempting to justify Univalence as a fundamental part of the foundations of mathematics, it should instead be viewed as an optional extra axiom, motivated on methodological grounds.

[2] The question of whether a mathematical structuralist should endorse the idea that "isomorphic objects are identical", and the broader question of the relationship between Univalence and mathematical structuralism suggested by Awodey in the quotations above, are the subject of a future paper and are not studied in detail here.

2 Universes

Informally, Univalence states that identity and equivalence are equivalent. More precisely it says that for two types `A` and `B`, the type expressing their identity is equivalent to the type expressing their equivalence. Of course, this is not intended to talk about just some specific two types `A` and `B` but rather is intended to be a general statement about arbitrary types. It therefore requires a way to quantify over types.

The basic rules of HoTT allow quantification over the tokens of an arbitrary type by the use of dependent pair types and dependent function types (which correspond roughly to existential and universal quantifiers). For example, given a predicate `P` defined on type `A`, a token of $\prod_{\mathtt{a:A}} \mathtt{P(a)}$ certifies that every token of `A` satisfies the predicate. Quantification over types therefore requires a type whose tokens are themselves types. Such a type is called a 'universe'.

While a complete understanding of universes involves many technical details that are beyond the scope of the present paper, Section 2.1 presents the core conceptual content of what's needed from universes in order to use them in HoTT and to understand Univalence. Section 2.2 briefly considers some of the technical details of implementation. Section 2.3 gives an account of universes that is compatible with the Types-as-Concepts interpretation of HoTT [Ladyman and Presnell, 2016] and explains how universes should be understood in an autonomous account of HoTT.

2.1 Rules for universes

Universes are introduced in [HoTT Book, Section 1.3] by the following facts:

- A universe $\mathcal{U}$ is a type whose elements are types.
- There is no universe $\mathcal{U}_\infty$ of all types.
- We have a hierarchy of universes $\mathcal{U}_0 : \mathcal{U}_1 : \mathcal{U}_2 : \ldots$
- Universes are cumulative: elements of $\mathcal{U}_i$ are also elements of $\mathcal{U}_{i+1}$.
- Every type (including every universe) inhabits some universe, and for any collection of types there is a universe containing them all.
- Universes are closed under the basic type-forming operations.
- A universe may be the domain or codomain of a function.

Before presenting an account of universes in an autonomous foundation, we first expand upon and explain the above.

Types in HoTT are ordinarily given in terms of token constructors and elimination rules. For example, the type of natural numbers $\mathbb{N}$ is defined in terms of constructors $\mathtt{z} : \mathbb{N}$ and $\mathtt{s} : \mathbb{N} \to \mathbb{N}$ which tell us what tokens of $\mathbb{N}$ can be produced, along with a means of defining functions that take natural numbers as input. Universes, although they are also types, are not presented in quite the same way. We are instead told that "A universe is

a type whose elements are types", and "When we say that $\mathtt{A}$ is a type, we mean that it inhabits some universe $\mathcal{U}_i$" [HoTT Book, Section 1.3]. There are not token constructors that produce the individual types that belong to a given universe. Rather, if anything plays this role for universes it is the basic rules of the language themselves, according to which for any universe $\mathcal{U}$ we have $\mathtt{0} : \mathcal{U}$ and $\mathtt{1} : \mathcal{U}$, and given $\mathtt{A}, \mathtt{B} : \mathcal{U}$ we have $\mathtt{A} \times \mathtt{B} : \mathcal{U}$, and so on.[3]

In practice, when defining a type we generally begin with some concept or proposition that we intend to capture in that definition – for example, the type of natural numbers, or the type corresponding to the statement of a particular theorem – and then work out how to express that idea in the language of HoTT. In contrast, a universe is not conceived of by starting with some defining idea that is to be expressed. Rather, the defining characteristic of a universe is, in general, just a particular collection of types that it is conceived of as having as its tokens. In this sense, whereas a type is conceived of *intensionally* as the realisation of a concept, a universe is conceived of as an *extensional* collection of types – those that are asserted to belong to the universe, along with those that can be constructed from them by application of the type formation rules of HoTT. [4]

In particular, if we conceive of a universe $\mathcal{U}$ as having types $\mathtt{A}, \mathtt{B}, \ldots$ amongst its tokens, then we must already have defined these types before defining $\mathcal{U}$ itself. Since $\mathcal{U}$ itself is not yet defined it cannot be one of these types, and so there cannot be a universe that contains itself – $\mathcal{U} : \mathcal{U}$ is not allowed.[5] As a consequence of this, there can be no 'universe of *all* types', since, being a type, this would have to contain itself as a token. In this way we avoid the paradoxes of type theory analogous to Russell's paradox in set theory.

Although no universe can contain itself, every universe, being a type, is contained in some universe. In particular we can define a universe that contains $\mathcal{U}$ along with all the types that are contained in $\mathcal{U}$. Moreover, we take it that whenever we have two universes $\mathcal{U}$ and $\mathcal{U}'$ such that $\mathcal{U} : \mathcal{U}'$ then $\mathcal{U}'$ contains all the types that are contained in $\mathcal{U}$; this principle is called **cumulativity**.

Note that this means that a type is a token of multiple universes, in contrast to the situation with ordinary types in which any token belongs to exactly one type. Indeed, any type belongs to potentially *infinitely many* universes: given any $\mathtt{A} : \mathcal{U}$, since $\mathcal{U}$ is a type we can define a universe $\mathcal{U}'$ containing it, and by cumulativity we have $\mathtt{A} : \mathcal{U}'$ as well; but then by the same argument we can introduce a further universe $\mathcal{U}''$ with $\mathcal{U}' : \mathcal{U}''$ and

[3] In the more formal presentation in [HoTT Book, Appendix] this is realised in the way the rules of the type theory are spelled out.

[4] We are grateful to a referee for this journal for pressing us to clarify this point.

[5] Indeed, if we allowed $\mathcal{U} : \mathcal{U}$ then the theory would be inconsistent. [Girard, 1972; Coquand, 1986]

$\mathtt{A} : \mathcal{U}''$, and so on (and since no universe can contain itself, each of $\mathcal{U}$, $\mathcal{U}'$, $\mathcal{U}''$, etc. must be different).

Since any universe is a type, it can be the input or output type of a function. This is one of the primary uses of universes. A predicate on some type $\mathtt{A}$ is a function taking a token $\mathtt{a} : \mathtt{A}$ as input and returning a type $\mathtt{P(a)}$ as output, i.e. a function of type $\mathtt{A} \to \mathcal{U}$. The fact that universes may be tokens of higher universes is essential to the expressive power of the theory. In order to discuss properties of some type $\mathtt{A} : \mathcal{U}$ we need predicates $\mathtt{P} : \mathtt{A} \to \mathcal{U}$, but the type $\mathtt{A} \to \mathcal{U}$ cannot be a token of $\mathcal{U}$. Thus in order to talk about the type to which such a predicate belongs we require a higher universe $\mathcal{U}'$. Similarly, to talk about properties of properties (e.g. satisfiability) we need higher-order predicates, and corresponding higher universes to which their types belong.

A universe may also be the domain of a function. This then allows us to define dependent functions that quantify over all types in a universe. For example, the token constructor of the identity type

$$\mathtt{refl} : \prod_{\mathtt{C}:\mathcal{U}} \prod_{\mathtt{x}:\mathtt{C}} \mathtt{Id}_{\mathtt{C}}(\mathtt{x}, \mathtt{x})$$

is intended to be read as saying 'all tokens of all types are self-identical'. While we can use a universe to quantify over 'all types' in this sense, we cannot quantify over all universes, since the domain of such a quantification would be a universe containing all universes as its tokens, and as we have said above no such universe exists. Thus we cannot translate a statement of the form 'for all universes $\mathcal{U}$, ...' into the language of HoTT.[6]

In general, the things we say about 'all types' within a given universe don't depend upon the particular features of that universe. For example, the function `refl` can be defined for all types in *any* universe – the existence of such a function isn't a special feature of some universes but not others. We can therefore understand these statements, definitions, and theorems as being implicitly extendable to any universe that is introduced. This is analogous to the situation in first-order logic, where we can't quantify over all predicates, but we regard first-order induction as a commitment to add an appropriate axiom of induction for any predicate that we introduce. In much the same way, instead of quantifying over all universes we can can treat '$\mathcal{U}$' as a dummy variable to be filled in with any particular universe, and whatever universe we choose will make the same theorems true (i.e. will allow us to carry out the same constructions). For many statements involving quantification over types – for example, the definition of `refl`, the token constructor for the identity type – this is the only reading that makes

[6] Similarly, for example, there is no function that takes an arbitrary universe $\mathcal{U}$ as input and returns as output a universe $\mathcal{U}'$ such that $\mathcal{U} : \mathcal{U}'$, since the input and output type of such a function would again be the (non-existent) type of all universes.

sense (in this case, that `refl` is defined for every type in every universe). Of course, our reasoning may involve more than one universe – as discussed above, we can have $\mathcal{U} : \mathcal{U}'$ and $\mathcal{U}' : \mathcal{U}''$ and so on. All these can likewise be treated as dummy variables, so long as we can fill in particular universes in such a way as to maintain the stated relationships between them. This informal technique of reasoning about an arbitrary universe $\mathcal{U}$ is called 'typical ambiguity', while the formal characteristic of the theory that enables this kind of reasoning is called 'universe polymorphism'. (For more on this see [HoTT Book, Section 1.3] and [Shulman, 2012].)

2.2 Some technical details

This section makes a few technical observations about universes, but a complete account is well beyond the scope of this paper.

One way to understand universes in HoTT is as a type-theoretic analogue of *Grothendieck universes* in set theory.[7] These are sets satisfying particular axioms that ensure that they're big enough to model ZFC (and thus the existence of a Grothendieck universe is not something that can be derived from the axioms of ZFC). Different approaches to the axiomatisation of Grothendieck universes place different constraints on how large such a set must be. In particular, some allow the empty set to qualify as a Grothendieck universe. Others demand that a universe may not be empty, but may be a countable set – for example, the set V_ω of hereditarily finite sets. Other approaches impose $\mathbb{N} \in \mathcal{U}$ as an axiom, which enforces that all universes be uncountable. In the approach taken here no universe can be empty (since we always have $0 : \mathcal{U}$ and $1 : \mathcal{U}$) but universes are not required to be uncountable.

Given the important differences between universes and other types noted above – in particular, that they have types as their tokens, that types belong to multiple universes rather than to exactly one, and that they are conceived of extensionally rather than intensionally – one might wonder whether universes should be thought of as types at all, rather than treating them as a different kind of thing altogether. There are indeed alternative approaches that treat universes in a different way.

For example, we might think of the tokens of universes as being 'codes' for types rather than types themselves, and then introduce an explicit 'decoding' function that produces for each code the corresponding type. This approach preserves more of the similarities between universes and other types, but is

[7] For a definition and discussion of Grothendieck universes, see [nLab, 2015]. To be clear, universes in HoTT are *not* Grothendieck universes, since the latter are defined set-theoretically, whereas set theory plays no role in HoTT. However, the analogy is sufficiently strong to be illuminating.

more cumbersome in practice.[8] Alternatively, one might take universes to be a third sort of thing (alongside tokens and types), and then modify the definition of functions to preserve the ability to use them as the domains and codomains of functions (which is an essential role that universes play in the theory).

Similarly, the assumption that the universes can be ordered as $\mathcal{U}_0 : \mathcal{U}_1 : \mathcal{U}_2 : \ldots$ can be dropped. This plays no essential role in the theory and may be regarded as a notational convenience, providing an easy way to produce names for universes. It is important to note that the indices used here are *not* tokens of the natural number type $\mathbb{N}$ in the theory, and in particular there is no function that maps a token $\mathtt{i} : \mathbb{N}$ to 'the corresponding universe' $\mathcal{U}_{\mathtt{i}}$ (not least because the output type of any such function is not defined).[9]

Each of these is a viable alternative to the account given above, and each has its own advantages and disadvantages. However, for simplicity (and for similarity with the approach taken in the HoTT Book) we adopt the above-described view of universes, and accept the fact that these must therefore be types of a rather different nature than we have previously considered. The next subsection gives some justification for these differences via the Types-as-Concepts interpretation.

2.3 Universes in an autonomous foundation

If HoTT is to be an autonomous foundation for mathematics then, whether or not Univalence is assumed, something must be said about how to understand universes and why the above rules governing their use are as they are. This subsection proposes an interpretation of universes that accounts for the properties outlined in Section 2.1 and is compatible with the Types-as-Concepts interpretation developed in [Ladyman and Presnell, 2016] which takes types to correspond to general mathematical concepts (such as 'natural number') and a token of a type to correspond to a specific mathematical concept *qua* instance of the more general concept (such as '2 *qua* natural number').

The Types-as-Concepts interpretation may be extended to universes by understanding them as *domains of discourse*, where a domain of discourse consists of the concepts and propositions that are understood and defined in a given discussion. This is a pre-mathematical notion – to have any discussion of any kind we must know what concepts and propositions are

[8] This is often called 'Tarski-style' universes, as opposed to the 'Russell-style' universes we have discussed (see [HoTT Book, Chapter 1 Notes]). For a more detailed comparison of the two approaches see [Luo, 2012].

[9] There are alternative proposals to augment the structure of universes by adding a 'super-universe' to which all universes belong, in which case a function of this kind could be definable. See [Palmgren, 1998; Dorais, 2014a;b].

within the domain – and is therefore admissible as part of an autonomous foundation. On this interpretation of universes we can explain and justify the characteristics of universes outlined above.

To consider the question of whether a given concept is part of a given domain of discourse it must be possible to conceive of the latter, and this is the concept associated with the type $\mathcal{U}$. There is nothing to a domain of discourse beyond what particular concepts it contains, and so domains of discourse are conceived of extensionally rather than intensionally (in the sense described above).

To give a complete characterisation of a mathematical concept we do not need to say anything about the domains of discourse that include it. Likewise in the type theory, as mentioned above types are in general defined without reference to any particular universe in which they may occur (except to the extent that the definition of a type depends upon other types or tokens, in which case any universe containing the type must also contain those types upon which it depends). Thus in this sense types are *primary*. Whereas the definition of tokens must follow the definition of the types to which they belong, the definition of universes must follow the definition of the types that they contain.[10] This explains why no universe can contain itself, and thus why there cannot be an ultimate domain of discourse that contains every concept we may ever need to consider.

Since it is possible to consider and discuss any concepts that can be precisely formulated, and any such concepts can be considered and discussed together, for any collection of concepts there is a domain of discourse that contains them all.

As argued in [Ladyman and Presnell, 2016; 2015b], the basic operations of HoTT correspond to the basic logical operations. Thus, since we take it that the discourse we are interested in is governed by logical rules, the domains of discourse under consideration are closed under the basic operations of HoTT. Thus, for example, if we can talk about some concept A and some concept B then we can also talk about their conjunction $A \,\&\, B$; more generally, if any concepts or propositions are part of our domain of discourse then so too must be anything that can be logically composed from them.

The definition of a token depends essentially upon the type to which the token belongs; we understand this in the Types-as-Concepts interpretation by saying that a token of some type corresponds to 'a particular concept *qua* instance of a general concept' – for example, '3 *qua* natural number' as opposed to '3 *qua* rational number'. We might summarise this by saying that any token has an essential intensional aspect that it derives from its

[10] Note that if the identity of a type did depend on what universe it was being considered as part of, then that universe (also being a type) would also have to be specified in terms of what universe it belongs to, and so on. This would lead to an infinite regress, and so nothing in the theory would be fully characterised.

type: it is not merely an object, but an object *thought about in a certain way*. This is used in [Ladyman and Presnell, 2016] to explain why each token must belong to exactly one type.

The relationship between types and universes, however, is different. As explained above, types are primary and do not depend for their definition upon any particular universe to which they might belong. Rather their definition is given by the type formation, token construction, and elimination and computation rules, none of which (in general) refer to any particular universe. Since the definition of, say, the natural numbers is given independently of and without reference to any universe, it makes no sense to think of '$\mathbb{N}$ *qua* element of universe $\mathcal{U}$' as something different from '$\mathbb{N}$ *qua* element of universe $\mathcal{V}$'. We may summarise this by saying that types do not pick up any further intensional character by being included in a particular universe. This explains why, in contrast to the usual situation with tokens and types, one type may belong to multiple universes.

We noted above that it is always possible to conceive of a particular domain of discourse itself, and that any precisely formulated concept belongs to some universe. Combining these observations with the fact that no domain can contain itself (because the contents of a domain must be defined before the domain itself can be), we see that each universe $\mathcal{U}$ must belong to some further (distinct) universe $\mathcal{U}'$. In other words, in conceiving of a particular domain of discourse we step outside that domain to a new one. Furthermore, since nothing is lost in this transition – no concept that had previously been defined is now no longer defined – the new domain contains all the concepts that were in the previous one. This justifies the assertion that domains of discourse are *cumulative* in the sense explained above.

A function between types `A` and `B` is something that, when given a token of `A`, produces a token of `B`. If `A` and `B` are thought of as propositions, with their tokens being 'proofs' or 'certificates', then this corresponds to material implication between those propositions. More generally, in the Types-as-Concepts interpretation, a function is something that takes an instance of one concept and produces an instance of the other. This understanding still holds when types `A` or `B` (or both) are universes. A function from a type `A` into a universe $\mathcal{U}$ is something that takes a token of `A` and returns as output a type belonging to that universe. In other words, when given a particular concept it produces a general concept whose definition may depend upon the given particular. In the other direction, a function from $\mathcal{U}$ to some type `B` takes as input a type from that universe and returns as output a token of `B`. In other words, it produces a particular instance of the general concept that may depend upon the general concept that is given as input. In this way, the use of universes as the domains and codomains of functions is just a natural generalisation of the way functions are defined on ordinary types.

In summary: Universes are types corresponding to domains of discourse

– at any point in our reasoning, the collection of types that we take to exist at that point is a universe. Universes have types as their tokens (and universes are the only types that have other types as their tokens), but no universe can be a token of itself. Thus there is no universe of *all* types. Universes are closed under all the basic rules of type formation. Thus there is no empty universe: the smallest universe must contain `0` and `1` and all the types arising from them via the basic rules.

As explained above, quantification over types involves specifying a universe. A 'univalent' universe is one in which Univalence holds of all the types in it. With typical ambiguity, explained above, the Univalence Axiom says that the universes we are working with are univalent.

The next section formally characterises Univalence.

3 The Definition of Univalence

Formally defining Univalence requires a definition of 'equivalence'. However, there is a subtlety to this which we postpone to Section 6. For now, think of 'equivalence' as a relation between types that is similar to isomorphism.[11] As a placeholder, we write $\mathtt{E}(\mathtt{A},\mathtt{B})$ for the type corresponding to this relation, whose tokens certify that types `A` and `B` are equivalent. For now we assume only that it is reflexive, that, like isomorphism, it is witnessed by the existence of a function $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ satisfying certain conditions, and in particular that the trivial function $1_{\mathtt{A}} : \mathtt{A} \to \mathtt{A}$ that leaves its input unchanged always counts as an equivalence.[12]

Recall that for any two tokens $\mathtt{x}, \mathtt{y}$ of a type `X` there is an identity type $\mathtt{Id}_{\mathtt{X}}(\mathtt{x},\mathtt{y})$, tokens of which are identifications of `x` and `y`. Since types themselves are tokens of any universe in which they live, there is an identity type $\mathtt{Id}_{\mathcal{U}}(\mathtt{A},\mathtt{B})$ for any pair of types, and its tokens are identifications between `A` and `B`.

Since the equivalence relation `E` is reflexive, there is a token of $\mathtt{E}(\mathtt{A},\mathtt{A})$ for any type `A`. It therefore immediately follows by path induction that for

[11] Indeed, to avoid complicating the informal presentation of [2014b], Awodey uses the word 'isomorphism' throughout much of the discussion and gives the standard category-theoretic definition. Note that in HoTT a relation is not in general merely a fact that either holds or fails to hold between two relata, but rather is represented by a family of types. We return to this in Section 6.2.

[12] It is possible instead to define an equivalence relation directly as a two-place predicate of type $\mathtt{A} \times \mathtt{B} \to \mathcal{U}$ satisfying certain conditions (see, for example, [HoTT Book, Exercise 4.2].) However, in general, rather than defining the equivalence relation $\mathtt{E}(\mathtt{A},\mathtt{B})$ directly we will instead define a corresponding predicate $\mathtt{isE}(\mathtt{f})$ on functions $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ asserting that `f` satisfies the relevant conditions. So, for example, rather than directly defining $\mathtt{Iso}(\mathtt{A},\mathtt{B})$ that says that `A` and `B` are isomorphic, we will define $\mathtt{isIso}(\mathtt{f})$ saying that `f` is an isomorphism between `A` and `B`.

any two types $\mathtt{A}$ and $\mathtt{B}$ there is a function

$$\mathtt{id\text{-}to\text{-}eq} : \mathtt{Id}_{\mathcal{U}}(\mathtt{A}, \mathtt{B}) \to \mathtt{E}(\mathtt{A}, \mathtt{B})$$

mapping identifications between $\mathtt{A}$ and $\mathtt{B}$ to equivalences between them. In other words, this function certifies that identical types are equivalent, as we would expect. In particular, this function maps the trivial self-identification $\mathtt{refl}_{\mathtt{A}}$ of type $\mathtt{A}$ to the trivial function $1_{\mathtt{A}} : \mathtt{A} \to \mathtt{A}$ (along with the proof that this trivial function is an equivalence).

Whereas the existence of the function `id-to-eq` follows from the basic rules of HoTT, the existence of a corresponding function from equivalence to identity does not. Moreover, in traditional mathematical frameworks it is possible to produce objects that are isomorphic but provably not identical, which therefore would block the introduction of such a function. In HoTT no such counterexamples can be produced. This leaves open the possibility of positing such a function.

When the Univalence axiom is added to the basic framework of HoTT it provides, for each pair of types $\mathtt{A}$ and $\mathtt{B}$, a function

$$\mathtt{eq\text{-}to\text{-}id} : \mathtt{E}(\mathtt{A}, \mathtt{B}) \to \mathtt{Id}_{\mathcal{U}}(\mathtt{A}, \mathtt{B})$$

and so under the assumption of Univalence, given any token of $\mathtt{E}(\mathtt{A}, \mathtt{B})$ witnessing the equivalence of $\mathtt{A}$ and $\mathtt{B}$, we can produce an identification between $\mathtt{A}$ and $\mathtt{B}$. Thus, under the assumption of Univalence, if we want to prove that two types are identical to one another it suffices to demonstrate that they are equivalent (in the particular sense denoted by '$\mathtt{E}$').

The existence of such a function `eq-to-id` is a radical departure from standard foundations. It allows us to set aside distinctions between isomorphic structures and treat them as presentations of a single structure in a fully rigorous way.

However, the Univalence axiom goes even further than this. Beyond merely asserting that a function `eq-to-id` exists, it says that this function and `id-to-eq` form an equivalence between $\mathtt{E}(\mathtt{A}, \mathtt{B})$ and $\mathtt{Id}_{\mathcal{U}}(\mathtt{A}, \mathtt{B})$ (again, in the sense of the relation denoted by '$\mathtt{E}$'). It therefore follows that the type of equivalences between $\mathtt{A}$ and $\mathtt{B}$ and the type of identifications between them are themselves equivalent:

$$\prod_{\mathtt{A},\mathtt{B}:\mathcal{U}} \mathtt{E}\big(\mathtt{Id}_{\mathcal{U}}(\mathtt{A}, \mathtt{B}),\ \mathtt{E}(\mathtt{A}, \mathtt{B})\big)$$

where we use quantification over the universe $\mathcal{U}$ to assert that this holds for all pairs of types $\mathtt{A}, \mathtt{B}$.[13]

[13] It is often said that this latter statement is the formal definition of Univalence. Strictly speaking this is not quite right: univalence says not just that there is *some* equivalence between

As noted above, in the basic framework of HoTT, Univalence cannot be proved as a theorem. That is, although the above type can be formulated in the language of HoTT, we cannot derive the existence of a token of that type. (Indeed, even the existence of a function of type $\mathtt{E(A,B)} \to \mathtt{Id}_{\mathcal{U}}\mathtt{(A,B)}$ cannot in general be proved.) Thus to use Univalence we must introduce it as an axiom by positing the existence of a token `UA` of this type.[14]

While the existence of a function `eq-to-id` is an important innovation, and to some extent already satisfies many of the aims of a structuralist approach to mathematics, it is important to note that the stronger assertion that $\mathtt{E(A,B)}$ and $\mathtt{Id}_{\mathcal{U}}\mathtt{(A,B)}$ are equivalent is an essential part of many applications of Univalence. We therefore introduce the terminology 'Semi-Univalence' for the axiom that just asserts the existence of a function $\mathtt{eq\text{-}to\text{-}id} : \mathtt{E(A,B)} \to \mathtt{Id}_{\mathcal{U}}\mathtt{(A,B)}$.

As Section 6 explains, like isomorphism, the equivalence relations `E` that are considered assert a strong correspondence between the tokens of the two types they relate: in particular, $\mathtt{E(X,Y)}$ entails that for each token of `X` there is a unique token of `Y`, and vice versa. Thus the connection asserted by Univalence between equivalence and identity allows any reasoning involving equivalences to be reduced to reasoning involving identifications without loss of information. In particular, if we want to prove that some property holds of all equivalences we can first re-express an arbitrary equivalence as $\mathtt{id\text{-}to\text{-}eq(p)}$ for some identification `p`, without loss of generality. This then reduces the problem to that of proving that some property holds of all identifications, and thus we can apply path induction to simplify the problem. In other words, the correspondence given by Univalence means that we can extend path induction to a corresponding statement about equivalences: to prove that a property holds of all equivalences it is sufficient to show that it holds of the trivial equivalence function $1_{\mathtt{A}}$ for all `A`. (This is the strategy involved in the proof of Theorem 3 in Section 6.5.)

Of course, to state the definition of Univalence properly we must fill in a particular equivalence relation for the placeholder '`E`'. For each such relation there is a corresponding variant of Univalence. In Section 6 we consider some possible definitions, and see that in some cases the resulting variant of Univalence is inconsistent, allowing us to derive contradiction.

$\mathtt{Id}_{\mathcal{U}}\mathtt{(A,B)}$ and $\mathtt{E(A,B)}$, but specifically that the particular function `id-to-eq` provides such an equivalence (with inverse given by `eq-to-id`). However, the technical reasons for making this distinction need not concern us and we will disregard it in the remainder of the paper. In Section 7.4 we address a different and more significant issue regarding the definition of Univalence.

[14] In computational implementations of Univalent HoTT there are disadvantages to having Univalence only as an axiom rather than a theorem. As noted in Section 1.1 here are therefore efforts to develop alternative versions of HoTT in which a version of Univalence can be proved as a theorem (for example, 'Cubical Type Theory' [Bezem et al., 2014]). However, such work is beyond the scope of the present paper, and so we refer to 'Univalence' and 'the Univalence axiom' interchangeably.

Before doing so, however, we first examine some of the consequences of adopting the Univalence axiom.

4 The Consequences of Univalence

4.1 Function Extensionality

An important consequence of Univalence is Function Extensionality (FE) which says that for any types $\mathtt{A}$ and $\mathtt{B}$, two functions $\mathtt{f}, \mathtt{g} : \mathtt{A} \to \mathtt{B}$ are equal iff they agree at all input values:

$$\left(\prod_{\mathtt{a}:\mathtt{A}} \mathtt{Id}_{\mathtt{B}}\big(\mathtt{f}(\mathtt{a}), \mathtt{g}(\mathtt{a})\big)\right) \;\to\; \mathtt{Id}_{\mathtt{A}\to\mathtt{B}}(\mathtt{f}, \mathtt{g})$$

In traditional mathematics this is taken as the identity criterion for functions, but it cannot be proved in basic HoTT. To make use of this important and useful principle, then, we must either assume FE directly as an axiom, derive it from another axiom, or extend the basic theory in some other way that gives FE.[15] Voevodsky's proof that FE can be derived as a theorem in HoTT under the assumption of the Univalence axiom (reconstructed in [Gambino et al., 2011]) makes essential use of the correspondence between equivalences and identifications given by Univalence. Semi-Univalence, which only asserts the existence of a function `eq-to-id` that maps equivalences to identifications, is not sufficient.

4.2 Non-trivial identifications

Another important consequence of UA is the existence of non-trivial self-identities.

HoTT is an intensional theory in the sense that tokens that are externally or 'judgementally' distinct may be nonetheless internally or 'propositionally' identical (meaning that the their identity type is inhabited).[16] However, the theory without Univalence admits extensional models in which all all identifications are trivial self-identifications. Put another way, without Univalence it is consistent to posit a 'reflection rule' saying that internal identity entails external identity. However, under Univalence we can directly prove the existence of a non-trivial self-identity, thus ruling out extensional models.

[15] One such approach is the addition of *Higher Inductive Types* [HoTT Book, Chapter 6] to the theory. These are types whose definition involves not only constructors for tokens but also constructors for identifications between tokens (and potentially higher identifications at all levels). If these are allowed then we may posit an 'interval type' consisting of two tokens and an identification between them. The existence of the interval type entails function extensionality. [HoTT Book, Lemma 6.3.2]

[16] For more on identity in HoTT, see [Ladyman and Presnell, 2015a].

Theorem 1. [HoTT Book, Example 3.1.9] Univalence entails the existence of a non-trivial identification.

Proof. Let $\mathtt{2} :\equiv \mathtt{1}+\mathtt{1}$, the coproduct of the Unit type with itself, whose tokens are $\langle *]$ and $[* \rangle$. The trivial function on $\mathtt{2}$ (that leaves its inputs unchanged) is mapped by Univalence to the trivial self-identity $\mathtt{refl}_2$. Define the function $\mathtt{swap} : \mathtt{2} \to \mathtt{2}$ by $\mathtt{swap}(\langle *]) :\equiv [* \rangle$ and $\mathtt{swap}([* \rangle) :\equiv \langle *]$. This is easily seen to be an equivalence, since it is self-inverse. Applying Univalence to $\mathtt{swap}$ gives an identification $\mathtt{p} : \mathtt{Id}_{\mathcal{U}}(\mathtt{2},\mathtt{2})$. Since $\mathtt{swap}$ and the trivial function are distinct, by Univalence the corresponding identifications must also be distinct, and so $\mathtt{p}$ is a token of $\mathtt{Id}_{\mathcal{U}}(\mathtt{2},\mathtt{2})$ that is not equal to $\mathtt{refl}_2$, i.e. a non-trivial self-identification of the type $\mathtt{2}$. ∎

Note that this proof makes essential use of the correspondence between equivalences and identifications given by Univalence – it is not sufficient merely to have a function `eq-to-id` that maps equivalences to identifications (as is given by Semi-Univalence), since this could not rule out the possibility that the identification produced was $\mathtt{refl}_2$.

In an extensional theory in which internal identity entails external identity, Univalence entails that all equivalent types are externally identical so there are no non-trivial equivalences. So although Univalence may be thought of as a kind of 'extensionality principle' in a certain sense (see Sections 5 and 8.3), is only of interest in an intensional system with distinct internal and external identities.

4.3 Other consequences

UA can be applied to the equivalence between identity types and equivalence types. That is, from the equivalence between $\mathtt{Id}_{\mathcal{U}}(\mathtt{A},\mathtt{B})$ and $\mathtt{E}(\mathtt{A},\mathtt{B})$ asserted by UA we can derive (by an application of UA) an identification of $\mathtt{Id}_{\mathcal{U}}(\mathtt{A},\mathtt{B})$ and $\mathtt{E}(\mathtt{A},\mathtt{B})$. This means that the transport function [HoTT Book, Lemma 2.3.1] associated with this identification can be used to translate facts about equivalences into facts about identifications and vice versa.

Without the Univalence axiom, while all functions respect identity, this does not extend to all predicates. It is true that for any function $\mathtt{k} : \mathtt{A} \to \mathtt{B}$, if we have $\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$ then we can derive $\mathtt{Id}_{\mathtt{B}}(\mathtt{f}(\mathtt{x}),\mathtt{f}(\mathtt{y}))$. However, the corresponding claim for predicates – i.e. given $\mathtt{P} : \mathtt{A} \to \mathcal{U}$, if $\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$ then we can derive $\mathtt{Id}_{\mathcal{U}}(\mathtt{P}(\mathtt{x}),\mathtt{P}(\mathtt{y}))$ – does not hold.[17] Without Univalence we can only prove under these circumstances that $\mathtt{P}(\mathtt{x})$ and $\mathtt{P}(\mathtt{y})$ are equivalent. The introduction of Univalence therefore puts predicates and functions on an even footing in this respect, since from this equivalence between $\mathtt{P}(\mathtt{x})$ and $\mathtt{P}(\mathtt{y})$

[17] See, for example, [Awodey, 2014a, 00:44:30], where this is described as a version of intensionality.

a corresponding identification can be derived. Thus, under Univalence, all predicates respect identity just as functions do.

The addition of Univalence imposes restrictions on what other axioms we can add to HoTT – in particular, how broadly the Law of Excluded Middle can be taken to apply. The basic rules of HoTT are grounded in constructive logic (as discussed in [Ladyman and Presnell, 2016]), but this does not constrain us always to use constructive methods because versions of the Law of Excluded Middle can be taken as additional axioms. However, there are limits to this: UA is inconsistent with the most general version of LEM that says that *every* type is either empty or inhabited:

$$\mathtt{LEM}_\infty :\equiv \prod_{\mathtt{A}:\mathcal{U}} \mathtt{A} + \neg\mathtt{A}$$

(The proof of this given in [HoTT Book, Corollary 3.2.7] makes use of the non-trivial self-identity of $\mathtt{2}$ derived in Theorem 1 above.) It is consistent with UA to assume a more restricted form of LEM that applies only to 'mere propositions' (i.e. those types having at most one token, up to identity; see [HoTT Book, Chapter 3.3] for further discussion of these types). Thus the use of Excluded Middle in the fragment of HoTT corresponding to propositional logic is still compatible with Univalence. (The situation regarding the Axiom of Choice is similar; see [HoTT Book, Section 3.8] for more details.)

Finally, an important consequence of Univalence with respect to the comparison between HoTT and other foundational systems is that HoTT with the Univalence axiom can recover standard ZFC set theory as well as Lawvere's Elementary Theory of the Category of Sets (ETCS). This entails that in so far as the whole of mathematics can be reconstructed in those theories, it can also be reconstructed in HoTT + UA. However, this is not to say that this is how HoTT is supposed to provide a foundation for mathematics in practice, rather this is done directly not by going via set theory [HoTT Book, Chapter 10].

5 The Meaning and Metaphysics of Univalence

While Univalence may be understood as saying that equivalence and identity are equivalent, and thus (as noted in Section 4) that they are identical, we must take care in interpreting this statement. Univalence does *not* say that the two notions are *externally* equal, and so even under Univalence we cannot replace identity by equivalence.[18] Given that the difference be-

[18] In particular, since equivalence is a relation between types, and no such notion is defined for tokens that are not themselves types, there is no way that identity could be systematically replaced by equivalence throughout the theory.

tween equivalence and identity is still recognised, in the sense that they remain externally distinct, what justifies their unification by UA? In this section we consider how to understand Univalence, and how it relates to our metaphysical picture of mathematics.

Univalence has been described as a kind of 'extensionality principle' for HoTT, since it unifies equivalent types that would otherwise be treated as distinct. However, note that, as mentioned in Section 4.2, in a type theory that is extensional in the sense that external identity is reflected in internal identity, Univalence trivialises equivalence. This is therefore not the sense in which Univalence adds a degree of extensionality to the theory. To see what it means to say that equivalent types are identical consider some examples:

(i) $\mathtt{A} \times \mathtt{B}$ and $\mathtt{B} \times \mathtt{A}$ are (externally) distinct types, but in almost any context in which we are interested in them there is no effective difference between them, so it makes sense to equate them.

(ii) The list-sorting algorithms MergeSort and InsertionSort are clearly distinct algorithms, and in some contexts the differences between them (e.g. their running times on a given list) are important. But in another sense, regarded just as relations between inputs and outputs, they are identical since they produce the same output when given the same input.

(iii) All empty types, for example, *even divisors of 9* and *largest prime*, are equivalent to $\mathtt{0}$ and thus (under Univalence) identical to it.

In each case the types that are equated by Univalence, being equivalent, were already *indistinguishable* within the language of HoTT, since any predicate that holds of one also holds of the other. However, while no predicate can discern two equivalent types this is not, in the basic theory, enough to enable us to prove that they are *identical*.[19] Thus while Univalence introduces new identifications between types that could not otherwise be produced, it only identifies types that were anyway already indiscernible. Moreover, recall that these are *internal* identifications, not external ones: the types involved remain externally distinct. In a sense, then, the external distinctions between types identified by Univalence record 'how the types were constructed or defined', which information is not accessible within the language itself.[20] So, for example, in the case of empty types the external distinctions between them record what particular impossible criteria and false propositions they correspond to, while internally the only fact about them that can be accessed is that they are empty. Since HoTT has these two

[19] As mentioned above, in [Ladyman and Presnell, 2015a] we investiate the relationship between identity and indiscernibility in HoTT, and in particular whether identity types might be re-interpreted as expressing indiscernibility. While this would naturally explain the various ways in which 'identity' in HoTT is unusual compared with standard theories, we argue that such a re-interpretation is not viable.

[20] We are grateful to Thorsten Altenkirch for pointing this out to us.

separate notions of identity we can still maintain the external distinctness of equivalent types and examine these concepts separately. Univalent HoTT therefore achieves a kind of identification without collapse, and so lets us keep both perspectives simultaneously.

Another interpretational issue regarding Univalence is how to understand the role of the universe $\mathcal{U}$. Any particular statement of the Univalence axiom in the language of HoTT involves quantification over all types in some particular universe $\mathcal{U}$. One way to read this is as a statement that equivalence is equivalent to identity for all types in *that particular universe* $\mathcal{U}$. A universe for which this holds is called **univalent** [HoTT Book, Section 2.10]. However, this is not the intended reading of the axiom. Rather than understanding it as an assertion about some *particular* universe $\mathcal{U}$, it is instead intended to be read in a 'typically ambiguous' way (Section 2.1) as an assertion about *all* universes, i.e. that all universes are univalent.

However, in the case of Univalence we may read the typical ambiguity in an alternative way, not as a metaphysical claim about universes, but rather as a methodological commitment to only consider univalent universes. This is quite compatible with there being non-univalent universes which we choose not to work with. Section 8 returns to this point.

5.1 Mathematical objects without Platonism?

The metaphysics of Univalence depends on the metaphysics of universes. The Types-as-Concepts interpretation given in Section 2 takes universes to be domains of discourse. On this interpretation, the two readings of Univalence given above correspond to the claim either that all domains of discourse are univalent, or the methodological commitment to restrict attention to univalent domains of discourse. Note that domains of discourse themselves may be interpreted in two different ways: as conceptual entities or as domains of objects. Platonists, who believe in the existence of mathematical objects, may take either interpretation (but might favour the latter), whereas non-Platonists are committed to the former interpretation.

In [1979, p. 32], Hersh records "a generally accepted fact about the mathematical world today: Most mathematicians live with two contradictory views on the nature and meaning of their work." Citing Hersh, Shapiro observes that

> "it is typical for a mathematician to be a Platonist during the week, when *doing* mathematics, and a formalist 'on Sunday', when there is leisure to think *about* mathematics. [...] This suggests that it is conducive to mathematics as such to treat, say, numbers *as if* they are part of an eternal, mind-independent realm, even if traditional Platonism, as an articulated philosophy, causes discomfort." [1997, pp. 28–9]

In this section we argue that Univalence may be seen as somewhat reconciling these two positions.

A Platonist can defend the Univalence axiom by appealing to the "eternal, mind-independent realm" of mathematical objects that they believe to exist. They can argue that equivalent types in the theory correspond to alternative descriptions of the same mathematical object (or the same collection of mathematical objects). Thus it makes sense to identify those types in the theory, in order to ensure that mathematical discourse is carved up at the right level of granularity, i.e. so that internally-distinct types correspond to different mathematical objects, not to different descriptions of objects.

However, Platonism throws up its own problems: for example, "the mathematician will wonder how it is possible to *know* anything about this eternal realm" [Shapiro, 1997, pp. 28–9]. Taken as a claim about mathematical ontology, then, Platonism is difficult to defend (see [Brown, 2011] for a recent defence of Platonism). Nonetheless, as Shapiro and Hersh observe, despite such philosophical objections mathematicians still conceive of mathematics in a Platonist manner, talking, writing, and (most importantly) thinking about mathematical entities as mind-independent objects.

This is where the Types-as-Concepts interpretation of HoTT brings an advantage. As discussed in [Ladyman and Presnell, 2016], this interpretation is silent about mathematical ontology, taking no position on the existence of the mathematical objects of which mathematicians treat. Rather, it takes the types of the theory to correspond to the *concepts* that are dealt with by mathematicians, without insisting that these concepts correspond to existing objects in the physical or Platonic realm. Our only ontological commitment is to the existence of concepts, to which we were anyway committed for non-mathematical reasons. Thus, on this interpretation, it is possible to accommodate mathematicians' Platonist treatment of mathematics without being committed to a Platonist ontology. The entities to which mathematical language appears to refer are taken to be concepts, whose existence we can comfortably assert without inflating our pre-existing ontology.

Moreover, on this interpretation we can employ the Platonist's argument for Univalence even without the corresponding ontology. Like the Platonist, we can argue that equivalent types in the theory correspond to a single mathematical concept under multiple different presentations. So, for example, it is natural to take "conjunction of `A` and `B`" and "conjunction of `B` and `A`" to be two different presentations of the same concept, and therefore to identify them in the theory.

Identifying types in HoTT lets us, in effect, create new types that are agglomerations of multiple finer-grained types. Whereas other systems, having a single identity relation, would require the addition of new entities to serve this function, HoTT does not need to do this. We can maintain the (external) distinctness of the original types, and so retain the intensional character

of the theory, while expressing their (internal) identity. So instead of having two classes of entities, namely 'objects' and 'names' (or 'references'), we have one class of entity and two ways of carving them up. One level of distinction is intensional, the other is extensional. Thus with Univalence we get the best of both worlds in an economical manner.[21]

However, while an argument along the above lines may indicate the appeal of a univalent theory and show how it fits naturally with mathematical practice and thought, there is more to be done. The account given in [Ladyman and Presnell, 2016] showed how the components of the basic language of HoTT – not including Univalence or function extensionality – could be motivated and justified from elementary pre-mathematical considerations, thereby defending the claim that HoTT may serve as an autonomous foundation for mathematics, not dependent upon any other foundational theory. If Univalent HoTT (i.e. the basic theory plus the Univalence axiom) is likewise to serve as an autonomous foundation then we must provide a pre-mathematical justification for the addition of the Univalence axiom. In particular, this must be an argument not just for the general idea of Univalence, such as the one sketched above, but for the exact Univalence axiom itself. So, before we can examine in detail the arguments for Univalence we must first fill in the remaining detail in its definition, namely the 'equivalence' relation denoted by '`E`' in Section 3.

6 Equivalence and Isomorphism

In standard mathematical parlance an equivalence relation is any relation that is reflexive, symmetric, and transitive. However, in most approaches to mathematical structuralism the relevant kind of equivalence is specifically the relation of isomorphism, while in category theory a weaker relation of equivalence between categories is appropriate, and in [HoTT Book, Chapter 4] a number of other equivalence relations are defined. This Section considers how the notion of equivalence should be formulated in HoTT, and how this affects the formulation of the Univalence axiom.

6.1 Isomorphism

According to the standard traditional definition of isomorphism, two structures A and B of some kind (for example, two groups, two manifolds, two

[21] We might try to get the same result without Univalence by using quotients. However, given that all equivalent types are being identified, it makes sense to do so in a uniform way by introducing a single axiom. Also, using quotients may involve constructing new types with counterparts and identifications, so introducing complications, whereas adding Univalence lets us keep the original types unchanged but adds identifications between them.

vector spaces) are isomorphic iff there exists a structure-preserving map $f : A \to B$ that has an inverse, i.e. a structure-preserving map $g : B \to A$ such that $g \circ f = id_A$ and $f \circ g = id_B$. (As noted in footnote 11, this is the definition that Awodey gives in [2014b, p. 9].)

The most straightforward translation of this into the language of HoTT is as follows: an isomorphism between types $\mathtt{A}$ and $\mathtt{B}$ is a function $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ having an inverse $\mathtt{g} : \mathtt{B} \to \mathtt{A}$ such that $\mathtt{g} \circ \mathtt{f} = \mathtt{id}_{\mathtt{A}}$ and $\mathtt{f} \circ \mathtt{g} = \mathtt{id}_{\mathtt{B}}$. Alternatively, if we wish to avoid using the axiom of function extensionality (which governs the identity conditions for functions, see Section 4.1) we may replace the conditions on $\mathtt{f}$ and $\mathtt{g}$ with

$$\mathtt{isPostInv_f(g)} :\equiv \prod_{\mathtt{a:A}} (\mathtt{g} \circ \mathtt{f})(\mathtt{a}) = \mathtt{a}$$
$$\mathtt{isPreInv_f(g)} :\equiv \prod_{\mathtt{b:B}} (\mathtt{f} \circ \mathtt{g})(\mathtt{b}) = \mathtt{b}$$

which say respectively that $\mathtt{g}$ is a post-inverse to $\mathtt{f}$ and a pre-inverse to $\mathtt{f}$.[22] We may then define

$$\mathtt{isIso(f)} :\equiv \sum_{\mathtt{g:B \to A}} \left(\mathtt{isPostInv_f(g)} \times \mathtt{isPreInv_f(g)}\right)$$

a token of which is a triple $(\mathtt{g}, \alpha, \beta) : \mathtt{isIso(f)}$, where $\mathtt{g} : \mathtt{B} \to \mathtt{A}$ is the inverse function and $\alpha : \mathtt{isPostInv_f(g)}$ and $\beta : \mathtt{isPreInv_f(g)}$ together certify that $\mathtt{g}$ is an inverse to $\mathtt{f}$.[23]

6.2 Mere propositions and truncation

Standardly in classical logic any proposition is merely either true or false. However, in HoTT we may interpret the types as corresponding to propositions, with tokens of types corresponding to certificates (or proofs) of those propositions. The possibility therefore arises that the type corresponding to a proposition may be inhabited by multiple distinct tokens, each of which certifies the truth of that proposition in different ways. We say that HoTT is a 'proof-relevant' system, in contrast to the 'proof-irrelevant' systems used in standard mathematics.

Some types, called 'mere propositions', behave more like the propositions of traditional proof-irrelevant logic, because they can be proved to have at most one token up to internal identity (i.e. any two tokens of the type are internally identical). Thus a token of a mere proposition provides just

[22] Post- and pre-inverse are more commonly called left- and right-inverses.

[23] In the HoTT Book (Definition 2.4.6) such a $\mathtt{g}$ is called a 'quasi-inverse' for $\mathtt{f}$, and the word 'isomorphism' is reserved for the case where types $\mathtt{A}$ and $\mathtt{B}$ satisfy a further condition (making them 'sets' or '0-types' in the terminology of the HoTT Book). We do not adopt that terminology in this paper.

the information that the proposition is true and no additional information. (Contrast with the proposition that some natural number n is composite, the proof of which might also provide us with a particular factor of n.)

In order to accommodate proof-irrelevant mathematics within the proof-relevant framework of HoTT, the HoTT Book introduces an operation of **propositional truncation** of a type [HoTT Book, Section 3.7].[24] Given a type $\mathtt{A}$, its truncation $\|\mathtt{A}\|$ is a type having a token for each token of $\mathtt{A}$, all of which are identified. Thus by definition it is a mere proposition that is inhabited iff $\mathtt{A}$ is. Taking the propositional truncation of a type instead of the type itself may be thought of as 'throwing away' information about which token of the type we may have and retaining only the fact of whether any such token exists at all, i.e. the truth or falsity of the corresponding proposition.

Since traditional approaches to mathematics have been proof-irrelevant by default (since no other option was available), when we come to translate ideas and definitions from standard mathematics into HoTT we sometimes have to choose between different possible formulations.[25] No clear guidance is given in the HoTT Book about when it is appropriate to use truncation in the translation of definitions into the language of HoTT. Thus any use of truncation in a translation should be given a specific justification, and we should not assume that definitions must be propositionally truncated (or, correspondingly, that types must be mere propositions) simply because the traditional definition necessarily was.

6.3 `isIso` is not a 'mere property'

Classically, *being an isomorphism* is merely a property that a given function $f : A \to B$ either has or does not have. But in HoTT, for a given function $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ there may be multiple distinct tokens of the type $\mathtt{isIso(f)}$, each certifying that $\mathtt{f}$ is an isomorphism.

We can rule out the possibility that $\mathtt{f}$ might have multiple inverses that behave differently, since if $\mathtt{g}, \mathtt{g'} : \mathtt{B} \to \mathtt{A}$ are both inverses to $\mathtt{f}$ then we have $\mathtt{g(b)} = \mathtt{(g \circ f \circ g')(b)} = \mathtt{g'(b)}$ for any $\mathtt{b} : \mathtt{B}$. But in general, for arbitrary $\mathtt{f}$, given an inverse $\mathtt{g}$ to $\mathtt{f}$ we cannot prove that all possible tokens $(\mathtt{g}, \alpha, \beta)$ and $(\mathtt{g}, \alpha', \beta')$ of $\mathtt{isIso(f)}$ are identical to one another. That is, we cannot

[24] The truncation operation cannot be defined in the basic language of HoTT without the addition of further axioms such as LEM [HoTT Book, Exercise 3.14] or the impredicative axiom of 'propositional resizing' [HoTT Book, Axiom 3.5.5, Exercise 3.15]. Alternatively it can be defined in HoTT extended with 'Higher Inductive Types', i.e. types whose constructors can produce identifications as well as tokens [HoTT Book, Section 6.9].

[25] This is similar to the situation we face when translating from classical to constructive mathematics, where distinctions that were collapsed by classical logic are now opened up, and we must choose which constructive variant best matches the original notion, or which best serves our purposes.

prove that for every function `f` the type `isIso(f)` is a 'mere proposition' [HoTT Book, Section 3.3].

Moreover, under the assumption of Univalence (and the presence of a propositional truncation operator) it is possible to construct a function `f` for which `isIso(f)` can be explicitly shown *not* to be a mere proposition [HoTT Book, Theorem 4.1.3]. In the terminology of [HoTT Book, Section 3.10], we might express this by saying that `isIso` fails to be a 'mere property' of functions.

In itself, this is not obviously a problem. While it conflicts with the assumption in standard mathematics that `f`'s *being an isomorphism* is a simple yes/no matter, we could take it as a discovery – obscured by previous proof-irrelevant frameworks – that being an isomorphism is a more complex property that may be witnessed in multiple ways. As noted in the previous section, we should not expect that every mathematical proposition that was previously assumed by default to be a mere matter of fact should turn out be so when translated into a richer proof-relevant system. Moreover we might hope to gain new insight by studying the structural relations between the various tokens of what was previously treated as a structureless mere fact.

In summary, then, there is initially no reason to think that there is anything wrong with this direct and immediate translation of the definition of isomorphism into HoTT (although we will see a very important reason in Section 6.5).

Before considering the consequences of using this definition of isomorphism we first consider two alternative equivalence relations that are 'mere properties', to see how they differ from isomorphism.

6.4 Truncated Isomorphism and Bi-invertibility

One way of ensuring that our translation of isomorphism into HoTT is a mere property is simply to put in this fact by hand by the use of a suitable truncation. If we define 'truncated isomorphism' $\|\texttt{isIso}\|$, where for each `f` we set $\|\texttt{isIso}\|(\texttt{f}) :\equiv \|\texttt{isIso(f)}\|$, then by definition a given function `f` can be a 'truncated isomorphism' in essentially one way. However there is no obvious motivation for introducing such a truncation.

The HoTT Book gives a number of alternative equivalence relations, all of which are mere properties in the sense defined in the previous section. In this section we examine the simplest such alternative, 'Bi-invertibility'.[26]

The definition of `isIso` corresponds to the standard definition of isomorphism, and requires proof that a single function `g` is *both* a pre-inverse and a post-inverse to `f`. This condition can be relaxed slightly, instead requiring

[26] For other examples, such as 'half adjoint equivalences' and 'contractible maps', see [HoTT Book, Chapter 4].

only that $\mathtt{f}$ has a pre-inverse and a post-inverse, without demanding in advance that the same function play both roles. This leads to an alternative to isomorphism, called 'bi-invertibility' [HoTT Book, Definition 4.3.1]

$$\mathtt{isBiInv(f)} :\equiv \left(\sum_{\mathtt{g}:\mathtt{B}\to\mathtt{A}} \mathtt{isPostInv_f(g)}\right) \times \left(\sum_{\mathtt{h}:\mathtt{B}\to\mathtt{A}} \mathtt{isPreInv_f(h)}\right)$$

with a corresponding relation between types

$$\mathtt{BiInv(A,B)} :\equiv \sum_{\mathtt{f}:\mathtt{A}\to\mathtt{B}} \mathtt{isBiInv(f)}$$

A certificate that $\mathtt{f}$ is bi-invertible is a quadruple $(\mathtt{g}, \alpha_{\mathtt{g}}, \mathtt{h}, \beta_{\mathtt{h}}) : \mathtt{isBiInv(f)}$, where $\mathtt{g},\mathtt{h} : \mathtt{B} \to \mathtt{A}$, $\alpha_{\mathtt{g}} : \mathtt{isPostInv_f(g)}$ and $\beta_{\mathtt{h}} : \mathtt{isPreInv_f(h)}$. Thus bi-invertibility requires only that a pre-inverse and a post-inverse to $\mathtt{f}$ exist, without requiring that they be the same function.

It is straightforward to show that this condition is materially equivalent to isomorphism:

Theorem 2. Let $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ be a function with a post-inverse $\mathtt{g} : \mathtt{B} \to \mathtt{A}$ and a pre-inverse $\mathtt{h} : \mathtt{B} \to \mathtt{A}$. It follows that (i) $\mathtt{g}$ is also a pre-inverse to $\mathtt{f}$; and (ii) $\mathtt{h}$ is also a post-inverse to $\mathtt{f}$.

Proof. Given $\mathtt{f},\mathtt{g},\mathtt{h}$ as above with $(\mathtt{g} \circ \mathtt{f})(\mathtt{a}) = \mathtt{a}$ for all $\mathtt{a} : \mathtt{A}$ and $(\mathtt{f} \circ \mathtt{h})(\mathtt{b}) = \mathtt{b}$ for all $\mathtt{b} : \mathtt{B}$ it follows that for any $\mathtt{b} : \mathtt{B}$, we have $(\mathtt{g} \circ \mathtt{f})(\mathtt{h}(\mathtt{b})) = \mathtt{h}(\mathtt{b})$ and, since all functions respect identity, $\mathtt{g}((\mathtt{f} \circ \mathtt{h})(\mathtt{b})) = \mathtt{g}(\mathtt{b})$. Thus $\mathtt{h}(\mathtt{b}) = (\mathtt{g} \circ \mathtt{f} \circ \mathtt{h})(\mathtt{b}) = \mathtt{g}(\mathtt{b})$. Under the assumption of function extensionality it would follow that $\mathtt{g} = \mathtt{h}$, which completes the proof; but even without this assumption we have (i) for any $\mathtt{b} : \mathtt{B}$, $(\mathtt{f} \circ \mathtt{g})(\mathtt{b}) = (\mathtt{f} \circ \mathtt{h})(\mathtt{b}) = \mathtt{b}$; and (ii) for any $\mathtt{a} : \mathtt{A}$, $(\mathtt{h} \circ \mathtt{f})(\mathtt{a}) = (\mathtt{g} \circ \mathtt{f})(\mathtt{a}) = \mathtt{a}$. ∎

The above theorem guarantees that the two functions $\mathtt{g}$ and $\mathtt{h}$ give the same values at all inputs, which is sufficient to show that they are both inverses of $\mathtt{f}$ (and under the assumption of function extensionality is sufficient to show that they are internally identical). Thus any bi-invertible function is also an isomorphism, and, trivially, any isomorphism with certificate $(\mathtt{g}, \alpha, \beta)$ is bi-invertible with certificate $(\mathtt{g}, \alpha, \mathtt{g}, \beta)$.

In a traditional proof-irrelevant approach to mathematics, the proof that $\mathtt{isBiInv(f)} \leftrightarrow \mathtt{isIso(f)}$ for all functions $\mathtt{f}$ would be taken to mean that for all purposes bi-invertibility and isomorphism are completely interchangeable. (This may explain why the notion of bi-invertibility never arises in standard mathematics, since in that proof-irrelevant context this modification of the definition of isomorphism is completely redundant.)

However, in the proof-relevant system of HoTT we can discern a difference between the two properties: using function extensionality it is possible to prove that $\mathtt{isBiInv}$ is a mere property [HoTT Book, Theorem 4.3.2].

Note that the distinction between `isIso` and `isBiInv` is a subtle one: it is not a matter of *which* functions have the property, since, as we have seen, any function having one property also has the other. Necessarily, then, the two criteria agree on which pairs of types count as 'equivalent'. Rather, the difference is a matter of *what evidence* must be presented to demonstrate that a function has the property: in the former case, a single inverse function bearing two properties; in the latter case, two 'one-sided' inverses, each bearing a single property. Moreover, as the above theorem shows, the two functions involved in a token of `isBiInv(f)` can be shown to be extensionally equal to one another, and thus each is a full inverse of `f` and could therefore serve as part of a token of `isIso(f)`. The distinction between these two properties is therefore not only subtle but also a deeply proof-relevant one. However, as we see in the next section, this distinction makes a very important difference for the formulation of Univalence.

6.5 'Isovalence' is inconsistent

Section 3 defined Univalence in terms of an as-yet unspecified relation of 'equivalence' that is "similar to isomorphism", and noted that for each particular equivalence relation there is a corresponding variant of Univalence. This section examines the result of defining Univalence in terms of the straightforward translation of isomorphism, `isIso` defined in Section 6.1. As we will be referring to this particular variant in subsequent sections, we introduce the name 'Isovalence'.

From `isIso` we define the corresponding relation between types

$$\mathtt{Iso}(\mathtt{A},\mathtt{B}) :\equiv \sum_{\mathtt{f}:\mathtt{A}\to\mathtt{B}} \mathtt{isIso}(\mathtt{f})$$

a token of which is a function $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ along with a token $(\mathtt{g}, \alpha, \beta) : \mathtt{isIso}(\mathtt{f})$. Thus $\mathtt{Iso}(\mathtt{A},\mathtt{B})$ may (very roughly) be understood as the type of all isomorphisms between `A` and `B` or, read as a proposition, as the assertion that such an isomorphism exists.

As in Section 3, it is straightforward to derive by path induction the existence of a function $\mathtt{id\text{-}to\text{-}iso} : \mathtt{Id}_{\mathcal{U}}(\mathtt{A},\mathtt{B}) \to \mathtt{Iso}(\mathtt{A},\mathtt{B})$ but in general the basic rules of HoTT do not allow us to produce a function in the opposite direction, $\mathtt{iso\text{-}to\text{-}id} : \mathtt{Iso}(\mathtt{A},\mathtt{B}) \to \mathtt{Id}_{\mathcal{U}}(\mathtt{A},\mathtt{B})$. Semi-Isovalence is the assertion that such a function exists, whereas Isovalence proper says that this function is an inverse to `id-to-iso`, and thus the types $\mathtt{Id}_{\mathcal{U}}(\mathtt{A},\mathtt{B})$ and $\mathtt{Iso}(\mathtt{A},\mathtt{B})$ are isomorphic to one another:

$$\prod_{\mathtt{A},\mathtt{B}:\mathcal{U}} \mathtt{Iso}\big(\mathtt{Id}_{\mathcal{U}}(\mathtt{A},\mathtt{B}),\ \mathtt{Iso}(\mathtt{A},\mathtt{B})\big)$$

If we took `isIso` to be a natural expression of equivalence, being the straightforward translation of isomorphism from traditional mathematics

into the language of HoTT, then correspondingly we should take the above statement to be the natural expression of the idea of Univalence, that 'equivalence is equivalent to identity'. However, this approach cannot be maintained as it leads to inconsistency.

For reasons of space we cannot give all the details of the proof here, but the outline of the proof is as follows. First, the derivation of Function Extensionality from Univalence (mentioned in Section 4.1) can be adapted to show that Function Extensionality also follows from Isovalence. Using this, it is then possible to reproduce the proof [HoTT Book, Theorem 4.3.2] that `isBiInv` is a mere property, and the construction given in [HoTT Book, Theorem 4.1.3] of a function `f` for which `isIso(f)` is not a mere proposition, thus demonstrating that `isIso` is not a mere property.

So far, this is just a reproduction of facts that hold under any version of Univalence. However, on the assumption of Isovalence we can also prove that `isIso` *is* a mere property, in contradiction to the above.[27]

Theorem 3. Let $\mathtt{A}, \mathtt{B}$ be any types and $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ any function between them. Then on the assumption of Isovalence, $\mathtt{isIso(f)}$ is a mere proposition.

To demonstrate this, we first prove a lemma. For a given function `f`, let `iso-b` be the obvious mapping that takes any token $(\mathtt{g}, \alpha, \beta) : \mathtt{isIso(f)}$ to $(\mathtt{g}, \alpha, \mathtt{g}, \beta) : \mathtt{isBiInv(f)}$, and let $\mathtt{b\text{-}iso} : \mathtt{isBiInv(f)} \to \mathtt{isIso(f)}$ be the mapping whose existence is proved in Theorem 2.

Lemma 1. Isovalence entails that for any function $\mathtt{f} : \mathtt{A} \to \mathtt{B}$, `b-iso` is a post-inverse of `iso-b`, i.e. for any $\mathtt{x} : \mathtt{isIso(f)}$ we have

$$\mathtt{b\text{-}iso(iso\text{-}b(x))} = \mathtt{x}$$

Proof. Isovalence says that every pair $(\mathtt{f}, \mathtt{x}) : \mathtt{Iso(A, B)}$ is the output of `id-to-iso` for some unique identification $\mathtt{q} : \mathtt{Id}_{\mathcal{U}}(\mathtt{A}, \mathtt{B})$. Thus we may assume without loss of generality that $(\mathtt{f}, \mathtt{x})$ is of the form $\mathtt{id\text{-}to\text{-}iso(q)}$. Then by path induction we may assume that $(\mathtt{f}, \mathtt{x})$ is $\mathtt{id\text{-}to\text{-}iso}(\mathtt{refl}_{\mathtt{A}})$, which is the trivial function $1_{\mathtt{A}} : \mathtt{A} \to \mathtt{A}$ that leaves its input unchanged, along with the obvious demonstration that this is an isomorphism. It is straightforward to show that the above equality holds in this case, and so by path induction and Isovalence it holds for any function `f` and any $\mathtt{x} : \mathtt{isIso(f)}$. ■

Theorem 3 now follows immediately:

Proof. For any tokens $\mathtt{x}, \mathtt{y} : \mathtt{isIso(f)}$ we have $\mathtt{iso\text{-}b(x)} = \mathtt{iso\text{-}b(y)}$, since $\mathtt{isBiInv(f)}$ is a mere proposition. Thus by Lemma 1 it follows that $\mathtt{x} = \mathtt{y}$, and so $\mathtt{isIso(f)}$ is a mere proposition. ■

Note that the crucial step in the above proof essentially involves the fact that Isovalence relates *isomorphisms* to identifications (specifically, it

[27] The proof sketched here follows that of [Gaillard, 2016].

allows us to reduce a statement about all isomorphisms to a statement about all identifications, to which we can then apply path induction). Thus no such inconsistency arises from a formulation of Univalence in terms of an equivalence relation such as `isBiInv` that is a mere property. Note also that the above proof of inconsistency makes essential use of Isovalence. 'Semi-Isovalence' (i.e. the existence of a function `iso-to-id`) is not inconsistent, and indeed follows from Univalence.

While it may seem entirely natural to define isomorphism in the standard way as `isIso` (as Awodey does in [2014b, Section 2]), and thus to read Univalence as saying 'isomorphism is isomorphic to identity', we now see that this reading leads to inconsistency.[28] Thus any attempt to justify Univalence as part of an autonomous foundation, along the lines discussed at the end of Section 5, must account for this by explaining why isomorphism as standardly defined, in the form of `isIso`, is not an acceptable formalisation of 'equivalence'. (In particular, the 'Platonist' argument sketched in Section 5.1 does not do this, since it simply takes the notion of 'equivalence' as given.)

In the HoTT Book [2013, Chapter 4] 'equivalence' is defined to be any mere property that is materially equivalent with `isIso`. Under Univalence, formulated in terms of any such relation, all these relations are equivalent to one another (although of course `isIso`, not being a mere property, is not) so the choice of which to use doesn't matter. Throughout this paper, unless otherwise indicated, we use 'Univalence' and 'UA' to denote such a consistent version of the axiom.

In Section 7.3 we consider reasons for using equivalence rather than isomorphism, as part of a discussion of an argument by Awodey that is the subject of the next section.

7 Univalence, Invariance, and Structuralism

In [2014b] Awodey gives an informal motivation for Univalence as an expression of the "Principle of [Mathematical] Structuralism", which he summarises as "isomorphic objects are identical".[29] Section 7.1 explains this informal argument in order to investigate whether it can be turned into an

[28] Although we asume that the authors were always aware of the inconsistency of Isovalence (which in [HoTT Book] is called `qinv-Univalence`), attention is not drawn to this fact: the proof is set as an exercise [HoTT Book, Exercise 4.6], which does not appear in editions of the book released before December 2013 [Bezem and Shulman, 2013].

[29] In the present paper we do not question whether this is a suitable summary of the central concept of mathematical structuralism, or a consequence of a structuralist approach to mathematics. We take up this issue in a future paper.

autonomous justification for Univalence. Sections 7.3 and 7.4 consider two problems that such an approach must overcome.

7.1 Awodey's argument for Univalence

Awodey begins with what he calls "the Principle of Structuralism", namely that "Isomorphic objects are identical" (and hence, *a fortiori*, have all the same properties) [2014b, p. 1]. This is "a principle of reasoning embodied in everyday mathematical practice" in the sense that "it makes no practical difference which of two 'isomorphic copies' are used, and so they can be treated as the same mathematical object for all practical purposes." However, as Benacerraf observed [1965] this principle is not upheld by standard set-theoretic foundations for mathematics. "Mathematical objects are often constructed out of other ones, and thus also have some residual structure resulting from that construction, in addition to whatever structure they may have as objects of interest" [Awodey, 2014b, p. 2]. We could instead try to find a system that upholds the weaker principle that 'Isomorphic objects have all the same properties', or at least that they share all their "relevant properties [...] pertaining to the subject matter" [Awodey, 2014b, p. 2]. This again is not upheld by set-theoretic foundations. This motivates a search for an alternative foundational system.

In order to develop this idea, 'isomorphism' must be defined. Awodey rejects a definition of 'isomorphism' as 'having the same structure', since this presumes that a definition of structure (or sameness of structure) is already available. Rather he gives the standard category-theoretic definition of isomorphism [2014b, Section 2] (which is the one stated in Section 6.1 above), and takes 'sameness of structure' to be defined as isomorphism. This justifies a definition of 'structural property' as any property that is invariant under isomorphism. Awodey then notes that in Martin-Löf's constructive type theory (on which the basic language of HoTT is based), "All definable properties are isomorphism invariant" [2014b, p. 6]. He calls this the "Principle of Invariance" (PI), which may be expressed formally as:

For any type $\mathtt{P}(\mathtt{X})$ definable over a basic type $\mathtt{X}$, $$\frac{\mathtt{A} \approx \mathtt{B} \qquad \mathtt{P}(\mathtt{A})}{\mathtt{P}(\mathtt{B})}$$

This can be proved to hold of all definable properties $\mathtt{P}$ whose definition does not involve any mention of the universe $\mathcal{U}$, but when the language is extended to include universes this proof no longer works.

Awodey argues that some further element should be added to the theory in order to extend this invariance principle to properties involving universes. With such an extended invariance principle we could take $\mathtt{P}(\mathtt{X}) :\equiv \mathtt{Id}_{\mathcal{U}}(\mathtt{A}, \mathtt{X})$, with the corresponding instance of PI:

$$\frac{\mathtt{A} \approx \mathtt{B} \qquad \mathtt{Id}_{\mathcal{U}}(\mathtt{A}, \mathtt{A})}{\mathtt{Id}_{\mathcal{U}}(\mathtt{A}, \mathtt{B})}$$

and thus derive from $\mathtt{A} \approx \mathtt{B}$ the conclusion $\mathtt{Id}_{\mathcal{U}}(\mathtt{A},\mathtt{B})$. Such an inference corresponds to the existence of a function from isomorphisms to identifications, i.e. a function `iso-to-id` as discussed in Section 3. Conversely, the existence of a function `iso-to-id` entails PI, because trivially reasoning is invariant under identity.

Thus if we want to build into our mathematical framework the idea that "Isomorphic objects are identical", and ensure that all our reasoning is invariant under isomorphism (thereby satisfying the Principle of Invariance), while allowing a language rich enough to encompass talk of universes, then Awodey's argument demonstrates that the existence of a function `iso-to-id` producing identifications from isomorphisms is necessary and sufficient.

7.2 The aim of Awodey's argument

To be clear, Awodey does not claim to be giving a formal justification for the Univalence axiom on the basis of the argument from Invariance. His argument is not intended to be a derivation or mathematical proof of Univalence from some more primitive assumptions. Rather, he asks whether it is possible to have an "extended system of type theory with a universe [in which] it is still the case that all definable properties are isomorphism invariant", noting that in such a system "isomorphic objects are identical" [2014b, p. 7]. He then notes [2014b, Section 5] that Univalent HoTT is indeed such a system, thus demonstrating that it is possible.

This observation that Univalence has, as one of its consequences, a principle of invariance that strongly accords with structuralist practice in mathematics, provides a justification for the adoption of the Univalence axiom. When we are choosing a mathematical foundation in which to work, several criteria come into consideration such as simplicity, ease of use, and correspondence with ordinary mathematical practice. None of the presently well-established foundational systems for mathematics, such as ZFC set theory or category theory, uphold the "Principle of Structuralism" that isomorphic objects are identical, nor the related "Invariance Principle" that all reasoning should be invariant under isomorphism. Awodey's argument demonstrates that a system containing the Univalence axiom does uphold these two principles. Thus when choosing amongst extant foundational systems, Awodey's argument shows that structuralist considerations motivate us to choose a system such as Univalent HoTT (or the Cubical Type Theory mentioned in Section 1.1).

However, the aim of our project is slightly different from Awodey's. [Ladyman and Presnell, 2016] gives a presentation of HoTT as an autonomous foundation for mathematics, showing how the components of the basic language of HoTT (not including universes or Univalence) could be motivated

and justified from elementary considerations. This section examines whether an argument along the lines of the one sketched above can be used as part of an autonomous presentation of Univalent HoTT. Our intention in the remainder of this section is not to criticise Awodey [2014b], but rather to attempt to clarify its details and extend it. The next two subsections identify two problems that must be overcome to do so.

7.3 The definition of isomorphism

As noted above, to avoid complication Awodey uses the word 'isomorphism' for the relevant relation involved in Univalence and gives the standard category-theoretic definition of isomorphism [2014b, p. 3]. While he notes that "Voevodsky's Univalence axiom itself actually has a more general form" involving a relation that is "a broad generalization of isomorphism" [2014b, p. 8]), he does not mention that the version of Univalence formulated in terms of isomorphism defined in this way (i.e. what we have called 'Isovalence') is inconsistent. However, of course, he does not intent his Argument from Invariance to be a justification of the inconsistent principle of Isovalence. Awodey and others who are familiar with HoTT are acutely aware of the difference between isomorphism and equivalence (as defined in Section 6) and the reasons for preferring the latter over the former. They therefore understand 'isomorphism' as a simplification for the more correct term 'equivalence', and appreciate why the substitution must be made. However, for more general readers not so familiar with HoTT this distinction and its importance may not be apparent. Our first aim in clarifying Awodey's argument in this section is therefore simply to point out this detail.

Having observed that such a substitution must be made, we must now proceed to justify it. For our purposes in this paper a proposed motivation for Univalence is only adequate if each step can be given an elementary pre-mathematical justification. To give an autonomous structuralist justification for Univalence along the lines of Awodey's argument we need a principled reason to think that 'sameness of structure' is captured best not by isomorphism (defined as `isIso`) but by a relation that makes Univalence consistent (i.e. something like truncated isomorphism or bi-invertibility that is a mere property). That is, the argument should begin as follows:

(i) The Principle of Structuralism says that mathematical objects that are *equivalent* in some suitable sense are taken to be identical;

(ii) By consideration of how we should understand mathematical objects from a structuralist perspective, the appropriate way to define the notion of 'equivalence' is as a mere property that is materially equivalent with `isIso`, such as 'truncated isomorphism' or 'bi-invertibility', whereas `isIso` itself fails to capture this notion correctly.

Such an argument needs to be spelled out in order to turn Awodey's infor-

mal motivation of Univalence into a more rigorous justification suitable for inclusion in an autonomous account of HoTT.

We therefore need to motivate the idea that the standard definition of isomorphism should be set aside and replaced with a relation such as truncated isomorphism or bi-invertibility in terms of which Univalence can be given a consistent definition. Note that, in doing so, it is not sufficient merely to justify taking such relations to be admissible alternatives to isomorphism: we must also justify why isomorphism itself (in its standard definition) should be regarded as inadmissible, or else Awodey's argument sketched above would still provide an equally good justification for Isovalence as for Univalence.

The authors of the HoTT Book say that `isIso` (which they call `qinv` for 'quasi-inverse') is "poorly behaved" [HoTT Book, Section 2.4] and "unsatisfactory because it is not a mere proposition, whereas we would rather that a given function could 'be an equivalence' in at most one way" [HoTT Book, Section 4.1]. But they do not present any specific argument for this preference.[30]

Of course, we could simply say that, since Univalence turns out to be such a useful principle, and since of course it must be formulated in a consistent way, we are therefore left with no option but to discard `isIso` and replace it with a relation such as `isBiInv`. While such a line of argument is mathematically reasonable – mathematicians, after all, are free to define their terms and choose their axioms in whatever way they find most useful – this is not a justification of Univalence along the lines of Awodey's argument, but rather a completely separate argument for Univalence in virtue of its usefulness (or its harmony with other mathematical ideas). Moreover, while arguments from mathematical usefulness or harmony have their place, they are not suitable as part of an autonomous presentation of a theory since they depend upon the sophisticated mathematical consequences that follow from the principle. Thus, while it may be reasonable to conclude that HoTT teaches us both that Univalence is a good idea and that isomorphism needs to be truncated or replaced as a notion of 'sameness of structure', this can't be the answer we're looking for.

Alternatively one might argue that, while notions such as material equivalence of propositions, bijection of sets, and isomorphism of set-theoretic structures are all appropriate notions of 'sameness' for their respective domains, the study of more sophisticated mathematical domains reveals that a more general notion of 'sameness' is required. In particular, in category theory and homotopy theory notions of 'weak equivalence' and 'homotopy equivalence' arise, which turn out to be a more suitable extension of the

[30] To be clear, this is not intended as a criticism of those authors since their aim is simply to present and explain the theory, not to motivate its various components from elementary pre-mathematical principles.

above notions. However, while this is an important insight (and moreover one that plays a crucial role in the unification of different domains at the heart of Homotopy Type Theory), this argument appeals to sophisticated mathematical developments and so cannot form part of an autonomous account of HoTT as a foundation.

A plausible elementary argument for equivalence and against isomorphism might proceed as follows. On a thoroughgoing structuralist view of mathematics all entities are to be individuated as structures, with none privileged as a separate category of entity with a distinct non-structural identity relation. Thus a purely structuralist notion of 'sameness of structure' should not be grounded in some other notion of 'sameness' such as identity, but should be free-standing. The definition of isomorphism, as stated in Section 6.1, involves the equations $g \circ f = id_A$ and $f \circ g = id_B$, and therefore fails this criterion. We might modify the definition of isomorphism to avoid this problem by replacing these two identities with some relation of 'sameness', but then this further notion of 'sameness' needs to be explicated. This threatens to lead either to a regress or a circularity. However, this can be avoided by giving a *coinductive* definition: any two mathematical entities A and B are '∞-isomorphic', $A \overset{\infty}{\approx} B$, iff there exist mappings $f : A \to B$ and $g : B \to A$ satisfying $g \circ f \overset{\infty}{\approx} id_A$ and $f \circ g \overset{\infty}{\approx} id_B$. While at first this appears to be circular, it turns out to be possible to give a consistent finite definition of such a relation.[31] Further, it can be proved that this relation is materially equivalent with `isIso` and is a mere property: that is, it is an equivalence in the sense of Section 6.

In summary, a structuralist justification for rejecting isomorphism and replacing it with an equivalence can indeed be given, as outlined above. In the proof-irrelevant frameworks in which mathematicians have worked up until recently this relation has been conflated with the materially equivalent relation isomorphism. Thus mathematicians have mistakenly taken isomorphism to be the natural criterion of 'sameness of structure', or have mistakenly referred to bi-invertibility by the name 'isomorphism', and it is only in a proof-relevant system such as HoTT that we can discern the difference and thus recognise the error.

To incorporate this into an autonomous account of a mathematical foundation we must address the objection that the notion of coinductive definition is itself mathematically sophisticated. This objection may be countered by, for example, more closely examining what is required of an autonomous foundation, and delineating between the motivating principles, which must be elementary and pre-mathematical, and the particular implementation, which involves mathematics. We defer more detailed examination of these ideas to another paper.

[31] Indeed, the notion defined in [HoTT Book, Section 4.2] as 'half-adjoint equivalence' is modelled on this idea.

In the remainder of this paper we set aside the above concerns and follow Awodey's usage of 'isomorphism' as a stand-in for the more correct relation of 'equivalence', a relation in terms of which Univalence can be formulated consistently. Even granting this, there is a more serious issue that remains to be settled.

7.4 Semi-Univalence vs Univalence

In Section 3 we distinguished between Semi-Univalence, which asserts the existence of a function `eq-to-id` from equivalences to identifications, and Univalence itself, which says that this function is an equivalence. Many important uses of Univalence, such as the derivation of Function Extensionality (Section 4.1), the existence of non-trivial identities (Section 4.2), and the proof that Isovalence is inconsistent (Section 6.5) require the strong correspondence between equivalences and identifications provided by Univalence, and cannot be derived using only Semi-Univalence.

However, the informal argument sketched above only motivates Semi-Univalence, not Univalence itself. If we are motivated by structuralist considerations to adopt a mathematical framework that validates the Principle of Invariance, Awodey's argument shows that it would be sufficient to add a function `eq-to-id`. As noted in Section 3, this is already a radical departure from standard mathematical foundations such as ZFC set theory, and is therefore in need of some kind of justification, which Awodey's argument provides. But his structuralist argument from Invariance does not in itself give us reason to assume further that such a function must be an equivalence, as Univalence states.

Awodey says that "Voevodsky's Univalence axiom itself actually has a more general form" than the claim that isomorphic types can be identified, namely that there is an equivalence between equivalence and identity [2014b, p. 9]. As we have argued, however, the difference between the Univalence axiom and the principle that is directly motivated by Awodey's informal argument is not a matter of "generalisation" but rather a considerable *strengthening*. Awodey's argument therefore provides a structuralist motivation for Univalence in the sense that it shows that desired consequences follow from that axiom, but does not give the complete justification that an autonomous presentation requires, since it does not give a principled reason for the adoption of Univalence as opposed to Semi-Univalence.

Of course, to some mathematicians it may seem obvious that once we've argued for a mapping that produces identifications from isomorphisms, then it is most natural to instantiate this in a *canonical* way, demanding that each identification correspond to a unique isomorphism and vice versa. But such an appeal to what seems natural to the intuitions of a mathematician cannot play a role in a autonomous justification of Univalence from pre-

mathematical principles.

Note, finally, that the 'Platonist' motivation for Univalence sketched in Section 5.1 suffers from the same flaw. While it motivates us to identify equivalent types by considering them as alternative presentations of a single underlying mathematical concept, this is only sufficient to justify Semi-Univalence.

An argument due to Dan Licata [Licata, 2016] (generalising from observations by Egbert Rijke and Martín Escardó [Rijke and Escardó, 2014] shows that the gap between Semi-Univalence and Univalence may be closed by the addition of a further principle that he calls $\mathtt{ua}\beta$. Recall that, given any identification between any two tokens $\mathtt{i} : \mathtt{Id_X(x,y)}$ and any predicate $\mathtt{P} : \mathtt{X} \to \mathcal{U}$, there is a transport function $\tau^{\mathtt{P}}_{\mathtt{i}} : \mathtt{P(x)} \to \mathtt{P(y)}$ [HoTT Book, Lemma 2.3.1]. In particular, given any identification between types $\mathtt{j} : \mathtt{Id}_{\mathcal{U}}(\mathtt{A},\mathtt{B})$, the transport of the trivial function $1_{\mathcal{U}} : \mathcal{U} \to \mathcal{U}$ gives a function $\tau^{1_{\mathcal{U}}}_{\mathtt{j}} : \mathtt{A} \to \mathtt{B}$. The principle $\mathtt{ua}\beta$ says that for any equivalence $\mathtt{f}$ (i.e. any function $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ for which there is a token $\mathtt{e} : \mathtt{isEquiv(f)}$), applying $\mathtt{eq\text{-}to\text{-}id}$ to this and then transporting $1_{\mathcal{U}}$ along the resulting identification produces the original function $\mathtt{f}$ itself, rather than some other function $\mathtt{A} \to \mathtt{B}$:

$$\mathtt{ua}\beta : \prod_{\mathtt{A},\mathtt{B}:\mathcal{U}} \prod_{(\mathtt{f},\mathtt{e}):\mathtt{Equiv}(\mathtt{A},\mathtt{B})} \tau^{1_{\mathcal{U}}}_{\mathtt{eq\text{-}to\text{-}id}(\mathtt{f},\mathtt{e})} = \mathtt{f}$$

From Semi-Univalence and this further principle, Univalence can be derived.[32] Thus if this further principle can be given an elementary justification then this would complete the autonomous presentation of Univalent HoTT.

One might argue for $\mathtt{ua}\beta$ as a 'conservation principle' along the following lines. Given only a function $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ that is known to be an equivalence, it should not be possible to use the given features of HoTT (namely Semi-Univalence and transport) to produce new functions $\mathtt{A} \to \mathtt{B}$. However, an argument based on this principle faces the problem that, even in the absence of $\mathtt{ua}\beta$, it is not possible to prove that $\tau^{1_{\mathcal{U}}}_{\mathtt{eq\text{-}to\text{-}id}(\mathtt{f},\mathtt{e})}$ and $\mathtt{f}$ are distinct (or else it would be inconsistent to posit $\mathtt{ua}\beta$). Thus, constructively, we have no reason to believe that what has been produced is a new function from $\mathtt{A}$ to $\mathtt{B}$; rather, it is just a function whose identity has not been established. However, the advocate of such a conservation principle may consider that this is already bad enough: if we have proved the existence of a function then it ought to be possible to know what particular function we have produced. Thus if we find the above conservation principle persuasive then this serves as motivation to adopt $\mathtt{ua}\beta$ in addition to Semi-Univalence. Moreover, the above argument does not appeal to sophisticated mathematical ideas beyond what is already present in basic HoTT, so it is compatible with an autonomous presentation of Univalent HoTT.

[32] Licata's proof, which he has formally verified in Agda, is given at the above citation.

8 Conclusion

8.1 An autonomous justification for Univalence?

This paper addresses the question whether Univalence can be given an elementary justification that can be part of a presentation of Univalent HoTT as an autonomous foundation for mathematics. It extends our Types-as-Concepts interpretation of HoTT to give an account of universes as domains of discourse, and argues that this captures and explains the features they have in the theory. It explains the definition of Univalence and in particular the definition of the equivalence relation that plays a central role in it, and why this cannot be naively understood as isomorphism. Finally, it examines Awodey's informal argument for Univalence from structuralist principles, identified two obstacles to developing it into an autonomous justification for Univalence, and outlined possible arguments in response to those issues. The adequacy of the resulting argument for Univalence therefore depends upon the answers to the following questions:

1. Is the account of universes as domains of discourse given in Section 2 adequate and autonomous (and if not, can some other autonomous account be given)?
2. Should we take a structuralist view of mathematics?
3. Does structuralism entail that "isomorphic objects are identical"?
4. Is the argument for ∞-isomorphism given in Section 7.3 autonomous, or does the reliance on coinduction to guarantee the non-circularity of the definition constitute an illegitimate appeal to mathematics?
5. Can the principle '$\mathtt{ua}\beta$' be given an autonomous justification, perhaps as a 'conservation principle', as outlined in Section 7.4?

Section 2.3 argued that 1 should be answered affirmatively. We postpone discussion of 2 and 3 to a future paper. Arguably, both questions 4 and 5 can be answered affirmatively, but the case for this is not as clear as for the definitions and rules involved in basic HoTT (as presented in [Ladyman and Presnell, 2016]). However, even if these arguments are found to be insufficient, the next subsection discusses an alternative reason to adopt Univalence.

8.2 Univalence as a methodological commitment

Recall from Section 5 that there are three possible readings of the statement of Univalence. Most straightforwardly, we might take it as the assertion that some particular universe $\mathcal{U}$ is univalent – but this is explicitly not the intended reading, since it is not possible to choose in advance a single universe in which all mathematical reasoning will take place. Instead we should take a typically ambiguous reading, viewing it either as the claim that

all universes are univalent, or as a commitment to only consider univalent universes.

If Univalence is understood as an assertion of mathematical fact that all universes are univalent then it is harder to defend. On the one hand there is no way within the theory to refer to, quantify over, or even conceive of 'all universes'; while on the other hand it is not part of the concept of a universe that it necessarily be univalent, since it is entirely possible to conceive of non-univalent universes. Likewise, under the Types-as-Concepts interpretation of universes as domains of discourse (Section 2) this claim would correspond to the assertion that all mathematical domains of discourse are univalent, and yet it is quite consistent to conceive of and work in a non-univalent domain.

Rather than defending the assertion that all universes are univalent, we might instead adopt Univalence as a methodological commitment to only consider univalent universes. Such an approach is much easier to motivate and to defend, since it is no longer required that the Univalence axiom in all its details be justified from pre-mathematical principles, and arguments of the type discussed in Section 7.2 become available. Given that we have reason to want an "extended system of type theory with a universe [in which] all definable properties are isomorphism invariant" [Awodey, 2014b, p. 7] the fact that Univalent HoTT provides such a system is reason enough to add the axiom. (Likewise, if one were attracted to the 'Platonist' argument in Section 5.1 then this too would provide sufficient reason to adopt Univalence in this way.)

What would be sacrificed on this approach, then, is the idea that Univalence is *true*, that it forms part of the foundation of mathematics offered by HoTT. Rather, it would be regarded as an optional additional axiom to be added to the basic rules of the foundation as required, just as, for example, the Law of Excluded Middle might be added in some particular applications.

This makes sense, since, despite the appeal of Univalence, we may sometimes have good reasons to work in basic HoTT (i.e. the theory without the addition of UA). For example, if we wanted to study a version of HoTT with the addition of a generalised form of the Law of Excluded Middle (applying to all types, not just mere propositions) this could not be done in the framework of Univalent HoTT, since this axiom is inconsistent with Univalence (Section 4.3). All the reasons we had to find basic HoTT interesting – for example, the parallels with homotopy theory and computer science – still apply and are not undermined by the existence of Univalent HoTT. The basic theory is therefore still worth studying even despite the greater strength of the Univalent theory.

8.3 The uses of Univalence

Regardless of the foundational status of the axiom, most of the use that is made of HoTT in practice assumed Univalence, which accords with the fact that the axiom is introduced early in [HoTT Book] and is then used throughout the remainder. Arguably the reason Univalence is useful is that it represents how mathematicians think by identifying types that would otherwise not be identified even though they are indiscernible. As explained in Section 4, Univalence allows us to exploit the benefits of the intensional nature of HoTT – viz. non-trivial identities and higher identity structure – whilst recovering the extensional character of much mathematical thought.

The language of HoTT without Univalence is so fine-grained as to distinguish types that differ only in how they are described, even if no property that can be defined in the language is able to distinguish them. Univalence allows us to identify such types, thus dispensing with such internally-imperceptible distinctions. The existence of two notions of identity – external and internal – lets us carve up types in a fine-grained way and a coarser-grained way. The existence of the fine-grained intensional view allows the formal language of HoTT to more faithfully reflect the way in which mathematical concepts are thought about and put together, while the coarse-graining blurs distinctions that are not mathematically important. Purely intensional theories such as basic HoTT are too fine-grained for ordinary mathematical practice in the sense that we do not have identifications between things that it is inconsistent to posit distinctions between. Purely extensional theories such as set theory collapses some of these distinctions, but introduce unwanted distinctions between different ways of representing mathematical structures (such as ordered pairs). By introducing an element of extensionality into the intentional theory of HoTT, Univalence strikes a balance between these two extremes.

References

Steve Awodey. Univalence as a new principle of logic. Video of a talk at the Calgary Mathematics & Philosophy Lecture Series, PIMS, University of Calgary, October 2014a.

Steve Awodey. Structuralism, Invariance, and Univalence. *Philosophia Mathematica*, 22(1):1–11, 2014b.

Steve Awodey, Álvaro Pelayo, and Michael A. Warren. Voevodsky's univalence axiom in homotopy type theory. *Notices of the AMS*, 60(9):1164–1167, 2013.

John C. Baez and James Dolan. Categorification. In Ezra Getzler and

Mikhail Kapranov, editors, *Higher Category Theory*, number 230 in Contemp. Math., pages 1–36. American Mathematical Society, Providence, Rhode Island, 1998.

Paul Benacerraf. What Numbers Could Not Be. *The Philosophical Review*, 74(1):47–73, 1965.

Marc Bezem and Michael Shulman. Why $isprop(isequiv(f))$ important? `https://github.com/HoTT/book/issues/553`, November 2013.

Marc Bezem, Thierry Coquand, and Simon Huber. A model of type theory in cubical sets. `http://www.cse.chalmers.se/~coquand/mod1.pdf`, 2014.

James Robert Brown. *Platonism, Naturalism, and Mathematical Knowledge*. Routledge, 2011.

T. Coquand. An analysis of Girard's paradox. In *Proceedings of the IEEE Symposium on Logic in Computer Science*, pages 227–236, 1986.

François Dorais. On the structure of universes in HoTT. 2014a. URL `http://logic.dorais.org/archives/1532`.

François Dorais. Super HoTT. 2014b. URL `http://logic.dorais.org/archives/1543`.

Pierre-Yves Gaillard. HoTT Reading Notes. "An informal set of comments on the HoTT Book", available at `http://iecl.univ-lorraine.fr/~Pierre-Yves.Gaillard/HoTT/ReadingNotes/HoTT_Reading_Notes_a.pdf`, February 2016.

Nicola Gambino, Chris Kapulkin, and Peter LeFanu Lumsdaine. The univalence axiom and functional extensionality. Notes prepared by Kapulkin and Lumsdaine from Gambino's lecture during the Oberwolfach Mini-Workshop on the Homotopy Interpretation of Constructive Type Theory (2011), 2011. URL `http://www-home.math.uwo.ca/~kkapulki/notes/UA_implies_FE.pdf`.

J.Y. Girard. *Interpretation fonctionelle et eleimination des coupures dans l'arithmetique d'ordre superieure*. PhD thesis, Université Paris 7, 1972.

Reuben Hersh. Some proposals for reviving the philosophy of mathematics. *Advances in Mathematics*, 31(1):31–50, 1979. doi: http://dx.doi.org/10.1016/0001-8708(79)90018-5.

James Ladyman and Stuart Presnell. A Primer on Homotopy Type Theory. Available at `http://www.bristol.ac.uk/homotopy-type-theory/`, 2014.

James Ladyman and Stuart Presnell. Identity in Homotopy Type Theory: Part II, The Conceptual and Philosophical Status of Identity in HoTT. under review, 2015a.

James Ladyman and Stuart Presnell. Identity in Homotopy Type Theory, Part I: The Justification of Path Induction. *Philosophia Mathematica*, 23 (3):386–406, 2015b.

James Ladyman and Stuart Presnell. Does Homotopy Type Theory Provide a Foundation for Mathematics? *British Journal for the Philosophy of Science*, 2016. Accepted for publication, available at `http://www.bristol.ac.uk/homotopy-type-theory/`.

Dan Licata. Weak univalence with "beta" implies full univalence. `https://groups.google.com/d/msg/homotopytypetheory/j2KBIvDw53s/YTDK4D0NFQAJ`, September 2016.

Zhaohui Luo. Notes on universes in type theory. 2012. URL `www.cs.rhul.ac.uk/home/zhaohui/universes.pdf`.

nLab. Grothendieck universe. 2015. URL `http://ncatlab.org/nlab/show/Grothendieck+universe`.

E. Palmgren. On universes in type theory. In *Twenty-five years of constructive type theory (Venice, 1995)*, number 36 in Oxford Logic Guides, pages 191–204. Oxford University Press, 1998.

Egbert Rijke and Martín Escardó. Generalize 7.2.2 and simplify encode-decode. https://github.com/HoTT/book/issues/718, October/December 2014.

Stewart Shapiro. *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press, 1997.

Michael Shulman. Universe polymorphism and typical ambiguity. Article at The n-Category Café, December 2012. URL `https://golem.ph.utexas.edu/category/2012/12/universe_polymorphism_and_typi.html`.

The Univalent Foundations Program. *Homotopy Type Theory: Univalent Foundations of Mathematics*. `http://homotopytypetheory.org/book`, Institute for Advanced Study, 2013.

Vladimir Voevodsky. Notes on type systems. Unpublished notes, 2009. URL `http://www.math.ias.edu/~vladimir/Site3/Univalent_Foundations.html`.