Philosophia Mathematica Advance Access published December 19, 2016

1–36
10.1093/philmat/nkw023
Philosophia Mathematica

Identity in Homotopy Type Theory: Part II, The Conceptual and Philosophical Status of Identity in HoTT†

James Ladyman* and Stuart Presnell**

*Department of Philosophy, University of Bristol, Bristol BS6 6JL, U.K.
E-mail: James.Ladyman@bristol.ac.uk
**Department of Philosophy, University of Bristol, Bristol BS6 6JL, U.K.
E-mail: Stuart.Presnell@bristol.ac.uk

ABSTRACT

Among the most interesting features of Homotopy Type Theory (HoTT) is the way it treats identity, which has various unusual characteristics. We examine the formal features of "identity types" in HoTT, and how they relate to its other features including intensionality, constructive logic, the interpretation of types as concepts, and the Univalence Axiom. The unusual behaviour of identity types might suggest that they be reinterpreted as representing *indiscernibility*. We explore this by defining indiscernibility in HoTT and examine its relationship with identity. We argue that identity types are a primitive component of HoTT and cannot be reduced to indiscernibility.

1. INTRODUCTION

One of the most interesting features of Homotopy Type Theory (HoTT) is the way it treats identity and the rich ∞-groupoid structure of types and

†We would like to thank two anonymous referees for their invaluable comments, observations, and suggestions, and Michael Shulman for detailed and very helpful feedback and discussion. The research on which this paper is based, 'Applying homotopy type theory in logic, metaphysics, and philosophy of physics', is funded by a Leverhulme Trust Research Project Grant (RPG-2013-228).

Philosophia Mathematica (III) Vol. 00 No. 0 © The Authors [2016]. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

identifications to which this gives rise.[1] This structure enables the homotopy interpretation of Martin-Löf's constructive intensional type theory that gives HoTT its name. Under this interpretation, types are interpreted as spaces (as understood in homotopy theory), tokens[2] of types as points of spaces, and identifications between tokens as *paths* or *homotopies* between points. Furthermore, the Univalence Axiom (due to Vladimir Voevodsky), which relates identity and equivalence, is taken by Steve Awodey [2014] to be fundamental to HoTT's claim to provide a new foundation for mathematics and to express mathematical structuralism.

This paper investigates the conceptual and philosophical status of identity in HoTT. We examine the formal and technical features of identity types in the theory in the context of its logic being constructive in Section 2, and how these relate to other features of the theory such as its intensionality and the interpretation of types as concepts in Section 3.[3] As we explain, in various respects the way identity is treated in HoTT departs considerably from orthodox views of identity in philosophy and logic. These departures are so radical that they lead us to question whether "identity" in HoTT is worthy of the name. Perhaps identity types should instead be interpreted as representing some other relation closely related to identity. Such an approach might offer a way of resolving many of the unusual features of "identity", and in particular might make Univalence seem very natural.

The axiom of Univalence is considered to be one of the main discoveries of the Homotopy Type Theory project. It roughly says that identity is equivalent to equivalence (where "equivalence" is a relation between types akin to isomorphism). As a consequence, it enables us to identify any two types that are isomorphic, thereby formalising as an axiom the idea that reasoning in mathematics should be invariant under sameness of structure. Yet, an important lesson of category theory is that we should 'never mistake an equivalence for an equality' [Baez and Dolan, 1998, p. 33], which seems, on a straightforward reading, to be in direct opposition to the Univalence Axiom. We begin to

[1] A comprehensive presentation of HoTT is given in the [The Univalent Foundations Program, 2013] (which we henceforth refer to as the "HoTT Book"). For a gentler introduction to the subject with more emphasis on philosophical issues, see [Ladyman and Presnell, 2014; 2015; forthcoming].

[2] Terminology varies in the literature — in the HoTT Book the words "term", "object", "element", and "point" are used interchangeably for what we are calling a "token" of a type. However, each of these words carries a connotation that we wish to avoid. In particular, "term" suggests something syntactic. We therefore prefer the word "token", which is (we hope) metaphysically neutral, and which we intend to be understood as a term of art independently of other usage.

[3] These are matters that are not made explicit in extant accounts which focus on the mathematics. In particular, while we draw heavily on the HoTT Book and some of what we say below makes explicit what is implicitly said there, our discussion of the philosophical interpretation of identity types and the majority of what we say about discernibility and indiscernibility is not found there in any form.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

explore the Univalence Axiom in Section 4.2 (see also [Ladyman and Presnell, forthcoming]).

In the standard presentation of HoTT, identity types are introduced as a primitive component of the system. Of course, in some logics identity is defined in terms of other resources, for example by the Hilbert-Bernays method [Caulton and Butterfield, 2012]. In Section 5 we motivate the idea that identity types might be better understood as encoding the *indiscernibility* of two tokens. In order to examine this option we consider how discernibility and indiscernibility might be formulated in HoTT in Section 6. This is of course closely related to the status of the Principle of the Identity of Indiscernibles (PII), and we examine different formulations of PII in a constructive setting in the Appendix. We consider finite types as potential counterexamples to PII in Section 7.1 and find that they do not threaten PII. However, independently of this question we find in Section 9 that treating identity as indiscernibility does not work. Hence, this paper provides a justification for why identity types are introduced as a primitive component of HoTT. Section 10 briefly concludes.

In this paper we assume a familiarity with the basics of HoTT, but in the next section we review in more detail the formal features of identity in HoTT.[4] (Note that unlike the presentation in HoTT Book we do not include the axioms of function extensionality and Univalence or the ability to define Higher Inductive Types as part of the basic language of HoTT.)

2. IDENTITY IN HoTT

As is standard in mathematics and logic, in HoTT identity is treated as a relation — for any two tokens of a given type, we may form the proposition saying that they are identical (which may or may not be true).[5] Since in HoTT propositions are represented by types in accordance with the Curry-Howard correspondence, the identity relation therefore takes a type $\mathtt{A}$ and two tokens $\mathtt{x}, \mathtt{y} : \mathtt{A}$ of that type and returns the **identity type** $\mathtt{Id}_{\mathtt{A}}(\mathtt{x}, \mathtt{y})$ corresponding to the proposition that tokens $\mathtt{x}$ and $\mathtt{y}$ are identical. Thus there is not a single identity type expressing all the identity facts about $\mathtt{A}$, but rather an identity type for each ordered pair of tokens of $\mathtt{A}$. Tokens of an identity type $\mathtt{Id}_{\mathtt{A}}(\mathtt{x}, \mathtt{y})$ are referred to as "identifications" of $\mathtt{x}$ and $\mathtt{y}$. The reflexivity of identity guarantees that the special case $\mathtt{Id}_{\mathtt{A}}(\mathtt{x}, \mathtt{x})$ has a token, for each $\mathtt{x} : \mathtt{A}$. Thus the token constructor for identity types produces a token $\mathtt{refl}_{\mathtt{x}} : \mathtt{Id}_{\mathtt{A}}(\mathtt{x}, \mathtt{x})$ for any token $\mathtt{x} : \mathtt{A}$. We sometimes use the standard notation "$\mathtt{x} = \mathtt{y}$" to indicate that the identity type $\mathtt{Id}_{\mathtt{A}}(\mathtt{x}, \mathtt{y})$ is inhabited.

Note that the proposition that two tokens are *not* identical corresponds to the negation of their identity type, $\mathtt{Id}_{\mathtt{A}}(\mathtt{x}, \mathtt{y}) \to \mathtt{0}$. (We sometimes use "diversity"

[4] See the HoTT Book or [Ladyman and Presnell, 2015] for a more detailed presentation of the basics of HoTT.

[5] Note that in standard HoTT only tokens of the same type may be identified — identity is "homogeneous". There are type theories, and in particular variants of HoTT, with "heterogeneous identity" (see for example [Licata and Brunerie, 2015]), but we do not consider these here.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

as a synonym for "non-identity" in this sense.) Thus, in particular, we cannot assert that two tokens of distinct types are non-identical (since there is no identity type between them of which we could form the negation). Also, since the logic of HoTT is constructive, some types may have tokens $\mathtt{x}$ and $\mathtt{y}$ such that neither $\mathtt{x} = \mathtt{y}$ nor $\mathtt{x} \neq \mathtt{y}$ can be proved (*i.e.*, identity for some types is not "decidable").

2.1. Internal versus External Identity

Fundamental to identity in HoTT is the distinction between "external" and "internal" identity, also described as the contrast between "judgmental equality" and "propositional identity".

Given any two expressions exp_1 and exp_2 naming tokens or types, the external identity $exp_1 \equiv exp_2$ says that exp_1 and exp_2 name the same token or type, and thus are intersubstitutable in any circumstance without changing meaning.

This is a different notion from the "internal" or "propositional" identity, which is represented by the identity types described above. In particular, internal identity, being represented by a type, can be combined with other types to make more complex propositions, such as conjunctions, disjunctions, and negations. Since external identity is not represented by a type in the theory corresponding to some proposition it cannot be combined in this way.

Note that external identity entails internal identity, *i.e.*, $\mathtt{x} \equiv \mathtt{y} \vdash \mathtt{Id_A(x,y)}$. This is because we always have a token of $\mathtt{Id_A(x,x)}$ for any token $\mathtt{x}$, and external identity enables substitution without change of meaning, so we may substitute $\mathtt{y}$ for $\mathtt{x}$ to conclude that we have a token of $\mathtt{Id_A(x,y)}$. However, in HoTT the converse implication does *not* hold: we may have internally identical tokens that are not externally identical. This is the definition of "intensional" for type theories [Martin-Löf, 1974]. A type theory having a "Reflection Rule" allowing external identity to be derived from internal identity is called "extensional".The fact that standard HoTT is intensional is very important as discussed in Section 3. We refer to externally non-identical tokens or types as "distinct".[6]

2.2. Multiple Identifications and Higher Identities

In standard approaches to mathematics the identity relation between two entities either holds or does not hold, and there is no more to be said. In HoTT, however, the identity relation between two tokens is a type, $\mathtt{Id_A(x,y)}$; if $\mathtt{x}$ and $\mathtt{y}$ are not identical then this type has no tokens, but if $\mathtt{Id_A(x,y)}$ is inhabited — *i.e.*, if $\mathtt{x}$ and $\mathtt{y}$ are identical — then this identity type may have *multiple* tokens. In other words, HoTT allows that two tokens $\mathtt{x}$ and $\mathtt{y}$ may be identified in multiple ways.

Note that in some contexts the idea of multiple identifications is quite natural. In particular, if one thinks of identifications as proofs of identity then the

[6] In contrast to internal identity, in standard HoTT any two tokens (of any types) can be compared for external identity, and all external identity facts are determinate and decidable.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

possibility of multiple identifications simply corresponds to the fact that there is often more than one way to prove that things are identical.

In the discussion of identity types at the start of this section we placed no restriction on the type $\mathtt{A}$ to which the tokens $\mathtt{x}$ and $\mathtt{y}$ belong. To capture the required expressive power it is essential that we are able to discuss the identity or non-identity of any two tokens of any type. In particular, this extends to the case where the tokens in question are themselves identifications of some $\mathtt{x}$ and $\mathtt{y}$. Thus, just as we can form the identity type $\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$ for any tokens $\mathtt{x}$ and $\mathtt{y}$ of any type $\mathtt{A}$, likewise given any two tokens α and β of $\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$ (*i.e.*, any two identifications of $\mathtt{x}$ and $\mathtt{y}$) we can form the so-called "higher identity type" $\mathtt{Id}_{\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})}(\alpha,\beta)$, corresponding to the proposition that these two identifications are themselves identical.[7]

This ramifies, so there is no limit in principle to the hierarchy of higher identity types: identifications of identifications may themselves be identified, and so on. It is this higher identity structure of multiple identifications at every level that gives rise to the ∞-groupoid structure explained in Section 2.4.

The possibility of multiple identifications and higher identities is fundamental to HoTT. The homotopy interpretation, which gives the theory so much of its interest, is based on thinking of tokens of types as points in spaces, and identifications between tokens as paths. Without multiple identifications the structure of identity types would be trivial, and so there could be no interesting connection between type theory and homotopy theory. In Section 3.2 we show that the intensional nature of HoTT is a necessary condition for there to be multiple identifications.

2.3. Path Induction

To use identifications we need the elimination rule for the identity type, called "path induction" (sometimes called "$\mathtt{J}$" by type theorists).[8] (We also need the associated computation rule; see [HoTT Book, Section 1.12.1] or [Ladyman and Presnell, 2015, Section 3.2] for details.)

For any type $\mathtt{A}$ and any token $\mathtt{a} : \mathtt{A}$ we can form the type

$$\mathtt{E}(\mathtt{a}) :\equiv \sum_{\mathtt{b}:\mathtt{A}} \mathtt{Id}_{\mathtt{A}}(\mathtt{a},\mathtt{b}),$$

whose tokens are pairs $\langle \mathtt{x}, \mathtt{i} \rangle$, where $\mathtt{x} : \mathtt{A}$ and $\mathtt{i} : \mathtt{Id}_{\mathtt{A}}(\mathtt{a},\mathtt{x})$. This can be thought of as the type of "all tokens of $\mathtt{A}$ that are identical to $\mathtt{a}$" (with some caveats, see [Ladyman and Presnell, 2015, Section 2.2] for details). In particular, one token of $\mathtt{E}(\mathtt{a})$ that is guaranteed to exist is $\langle \mathtt{a}, \mathtt{refl}_{\mathtt{a}} \rangle$.

[7] Extending the analogy with proofs, there may be proofs α and β that are written differently but are in fact identical. For example, perhaps they differ only in the order in which two commuting operations are performed. In this case we would have a "higher order" proof that the proofs α and β are identical.

[8] Strictly speaking what we present here is a slightly simpler variant called "based path induction", which is equivalent to path induction [HoTT Book, Section 1.12.2].

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

Based path induction says that for any type $\mathtt{A}$, any token $\mathtt{a} : \mathtt{A}$, and any predicate $\mathtt{K}$ that can be asserted of tokens of $\mathtt{E(a)}$, if $\mathtt{K}$ holds of $\langle \mathtt{a}, \mathtt{refl_a} \rangle$ then it holds of *all* tokens $\langle \mathtt{x}, \mathtt{i} \rangle : \mathtt{E(a)}$. More formally, there is a function

$$\mathtt{K}(\langle \mathtt{a}, \mathtt{refl_a} \rangle) \rightarrow \prod_{\langle \mathtt{x},\mathtt{i} \rangle : \mathtt{E(a)}} \mathtt{K}(\langle \mathtt{x}, \mathtt{i} \rangle).$$

Path induction may be interpreted as saying that identity is the "least reflexive relation" ("least" here in the sense of minimal or initial) [Martin-Löf, 1974]. Alternatively it may be understood in terms of retractions of paths in the homotopy interpretation [HoTT Book, Remark 1.12.1]. In [Ladyman and Presnell, 2015] we discuss its epistemological and methodological status, and give an alternative elementary justification for this rule that does not depend upon pre-existing mathematical notions.[9]

2.4. The ∞-Groupoid Structure of Identity

Using path induction it is straightforward to prove that the expected features of identity, such as symmetry and transitivity, hold. More specifically, given any identification $\mathtt{i} : \mathtt{Id_A(x, y)}$, we can derive a corresponding identification $\mathtt{i}^{-1} : \mathtt{Id_A(y, x)}$, and similarly given identifications $\mathtt{i} : \mathtt{Id_A(x, y)}$ and $\mathtt{j} : \mathtt{Id_A(y, z)}$ we can derive an identification $\mathtt{j} \,\square\, \mathtt{i} : \mathtt{Id_A(x, z)}$ (where we write " $\square$ " for the composition operation for identifications).

Given that these relations hold, we might try to define a category whose objects are the tokens of some given type $\mathtt{A}$ and whose arrows are the identifications between those tokens. For each object $\mathtt{x}$ in the category the self-identification $\mathtt{refl_x}$ gives the identity arrow, while transitivity of identifications shows that arrows can be composed. Moreover, since identifications are symmetric it follows that this category, if it existed, would be a **groupoid** (*i.e.*, a category in which every arrow has an inverse).

However, there is a problem with this definition. Composition of arrows in a category must satisfy certain requirements, such as associativity. However, the composition of identifications is not strictly associative: given three identifications, the respective compositions are not externally equal,

$$(\mathtt{k} \,\square\, \mathtt{j}) \,\square\, \mathtt{i} \;\not\equiv\; \mathtt{k} \,\square\, (\mathtt{j} \,\square\, \mathtt{i})$$

(*i.e.*, they are not literally the same identification), but rather there is an internal identification between them (*i.e.*, a token of the corresponding higher identity type). Likewise for the other requirements on arrows: in each case, rather than satisfying the required equation strictly by an external identity

[9] See [Ladyman and Presnell, 2016] for further discussion of what is required of a foundation for mathematics (in the sense that a philosopher understands the term "foundation"), and for arguments why one might object to a purported foundation whose basic rules are explained or motivated via sophisticated mathematical ideas such as homotopy theory.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

there is instead an internal identification. Thus the structure arising from the tokens and identifications in a type is not a groupoid, but rather an ∞-**groupoid** [HoTT Book, Section 2.1].[10]

This ∞-groupoid structure is central to the interest of HoTT. It was known before HoTT that there is a correspondence between ∞-groupoids and homotopy theory. It is this structure of the identity types that enables the homotopy interpretation of the type theory.

Having reviewed the definition and novel features of identity in HoTT, in the next section we reflect upon them.

3. THE CONCEPTUAL AND PHILOSOPHICAL STATUS OF IDENTITY IN HoTT

Identity types and their structure can be regarded as formal features of HoTT that are of mathematical interest independently of their conceptual and philosophical status. However, the latter is also novel and interesting: indeed, the features of the identity type in HoTT are so unusual that in Section 3.4 we consider whether it deserves the name "identity" at all. Before coming to that, we explicate some of its features.

3.1. Absolute or Relative?

In many systems we can ask whether any two arbitrary elements are identical to each other. For example, in ZFC, we can always ask whether a set is equal to another, whether an element of one set is equal to an element of some other set, and so on. Moreover, in ZFC identity is **absolute**: there is a single identity relation in which entities may or may not stand, and this is not relativised in any way. This is in contrast with the notion of **relative** identity [Geach, 1967], whereby two entities may be identical in one respect but distinct in another. For example, if a figure F_1 is sculpted from clay, and this piece of clay is then deformed and re-sculpted into a new shape F_2, we might say that F_1 and F_2 are the same *qua* piece of clay but distinct *qua* statue. In this case the identity or non-identity of F_1 and F_2 is not an absolute matter, but relative to the *intension* under which they are considered.

As discussed in Section 2, the identity relation in HoTT is homogeneous, meaning that only tokens of the same type may be identified. We might therefore initially guess that the distinction between external and internal identity is exactly the difference between absolute and relative identity. That is, since $\mathtt{x} \equiv \mathtt{y}$ allows us to substitute $\mathtt{x}$ and $\mathtt{y}$ for each other freely in any context without change of meaning, we might try to interpret $\mathtt{x} \equiv \mathtt{y}$ as the absolute identity of $\mathtt{x}$ and $\mathtt{y}$. Internal identity, on the other hand, appears to be relative in the above sense: a token of $\mathtt{Id}_{\mathtt{A}}(\mathtt{x}, \mathtt{y})$ demonstrates that $\mathtt{x}$ and $\mathtt{y}$ are the same *qua* token of type $\mathtt{A}$. (This view might seem to follow from the Types-as-Concepts interpretation presented in [Ladyman and Presnell, 2016], in which types are

[10] Although it is not generically the case that all types are groupoids, one can define a model of (the basic language of) HoTT in which this holds [Hofmann and Streicher, 1998].

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

general mathematical concepts and tokens are specific mathematical concepts regarded as instances of the more general one. For example, the type of *metric spaces* has as tokens particular metric spaces such as the Euclidean plane.)

However, in HoTT every token belongs *essentially* to a particular type, and cannot belong to two distinct types. Thus it is not possible for two tokens to be identical in one type but non-identical in another, since this would require that a token belong to more than one type. Thus in HoTT internal identity is absolute, in the sense that the identity of tokens cannot vary across contexts or aspects.

Recall from Section 2 that if `a` and `b` are of different types then we also cannot assert that they are *non-identical*, since this proposition corresponds to the negation of an identity type between `a` and `b`, and the latter does not exist. For example, since the type of natural numbers is distinct from the type of real numbers, there is no question as to whether the natural number 2 is identical to the real number 2. (This is in roughly in accordance with how identity is treated in Michael Resnik's version of mathematical structuralism [1997].)[11]

3.2. Intensionality

The intensionality of HoTT, in the sense of there being two identity relations (external and internal) that come apart, is an important source of its expressive power.[12] It enables us to make fine-grained (external) distinctions that bring greater conceptual clarity while at the same time allowing us to identify (internally) tokens that are the same in an appropriate sense. This makes it possible to work at just the right level of granularity for a given application. Recall that this is only possible because the two notions of identity are distinct: while external identity entails internal identity (Section 2.1), the converse implication from internal to external identity does not hold.[13]

The Types-as-Concepts view introduced in [Ladyman and Presnell, 2016] provides a natural way to understand the intensionality of HoTT (*i.e.*, the fact that it is posible to have tokens that are externally distinct but internally identified). For example, consider two distinct Euclidean planes in Euclidean 3-space; these are two distinct tokens of the type of *metric spaces*, but arguably

[11] There is a way to represent the relationship between the natural number 5 and the prime number 5 using dependent types. For more details, see [Ladyman and Presnell, 2014] and the HoTT Book. When working formally with an automated proof assistant the same symbol can be used for a natural number and the corresponding real number since the inclusion of $\mathbb{N} \to \mathbb{R}$ is taken to be an "implicit coercion". We are grateful to an anonymous referee for this journal for pointing this out.

[12] Note that this intensionality is a feature of the formal system, and is independent of how we interpret the "internal identity" relation or what we call it. Thus the proposals considered below to reinterpret $\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$ as denoting indiscernibility instead of identity, or to replace it with an alternatively-defined type that satisfies the same properties as $\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$, do not affect the intensionality of the theory.

[13] Voevodsky has investigated ways to define a language with two internal identity relations, one extensional and the other intensional [Voevodsky, 2013; nLab, 2014], and it is possible to have two such relations in a language if suitable constraints are imposed.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

we should think that they are nonetheless identical *qua* metric space. The intensionality of HoTT means that we can use distinct tokens to capture the fact that we are thinking about the two copies of the Euclidean plane differently, even though thought of *qua* metric space we identify them.

For another example, consider the hyperbolic disc and the Euclidean plane as metric spaces. These are distinct *qua* metric space, but if we forget about their metric structure then the topological spaces we obtain are the same *qua* topological space. Extensionally, "the underlying topological space of the hyperbolic disc" and "the underlying topological space of the Euclidean plane" are two names for a single mathematical object. However, it may be useful to record the fact that we are thinking about this topological space in two different ways. In an intensional theory like HoTT we can do so, representing them by two extensionally distinct tokens, which may be internally identified in accordance with the fact that they are identical *qua* topological space.

3.3. Intensionality and Path Induction

It may seem that path induction (Section 2.3) is in conflict with the intensionality of the theory. Recall that this rule says that if some predicate $\mathtt{K}$ holds of $\langle \mathtt{a}, \mathtt{refl}_{\mathtt{a}} \rangle : \mathtt{E}(\mathtt{a})$ then it holds of all tokens $\langle \mathtt{x}, \mathtt{i} \rangle : \mathtt{E}(\mathtt{a})$. A simple way to interpret this would be to say that $\langle \mathtt{a}, \mathtt{refl}_{\mathtt{a}} \rangle$ is the *only* token of $\mathtt{E}(\mathtt{a})$, and thus the only token identical with $\mathtt{a}$ is $\mathtt{a}$ itself, and only trivially so due to the reflexivity of identity.

However, path induction does *not* allow us to prove that all identifications are trivial. To prove this we would need a predicate $\mathtt{Q}$ that says of any given identification $\mathtt{p} : \mathtt{Id}_{\mathtt{A}}(\mathtt{a}, \mathtt{b})$ that it is identical to some trivial self-identification $\mathtt{refl}_{\mathtt{c}}$ of some token $\mathtt{c} : \mathtt{A}$. However, such a trivial self-identification is a token of $\mathtt{Id}_{\mathtt{A}}(\mathtt{c}, \mathtt{c})$, whereas $\mathtt{p}$ is a token of $\mathtt{Id}_{\mathtt{A}}(\mathtt{a}, \mathtt{b})$. While external identity allows arbitrary substitution in any context, internal identity does not. Thus the internal identification of $\mathtt{a}$ and $\mathtt{b}$ does not entail that $\mathtt{p}$ is a token of $\mathtt{Id}_{\mathtt{A}}(\mathtt{a}, \mathtt{a})$ (or of $\mathtt{Id}_{\mathtt{A}}(\mathtt{b}, \mathtt{b})$). Thus there is no type that has both $\mathtt{p}$ and a trivial self-identification as tokens. Since we can only express the identity of tokens belonging to the same type, the claim that an arbitrary identification $\mathtt{p}$ is trivial would not be well-typed, and therefore cannot be expressed in the language of HoTT, let alone proved.

However, note that this depends upon the intensionality of the theory, and in particular on the absence of a Reflection Rule deriving external identity from internal identity (see [HoTT Book, Exercise 2.14]). If the theory were extensional then from $\mathtt{p} : \mathtt{Id}_{\mathtt{A}}(\mathtt{a}, \mathtt{b})$ we could derive $\mathtt{a} \equiv \mathtt{b}$, and then by substitution we would have $\mathtt{p} : \mathtt{Id}_{\mathtt{A}}(\mathtt{a}, \mathtt{a})$. Then we could define the predicate

$$\mathtt{Q}(\langle \mathtt{b}, \mathtt{p} \rangle) :\equiv \mathtt{Id}_{\mathtt{Id}_{\mathtt{A}}(\mathtt{a},\mathtt{a})}(\mathtt{p}, \mathtt{refl}_{\mathtt{a}}).$$

This (trivially) holds for $\langle \mathtt{a}, \mathtt{refl}_{\mathtt{a}} \rangle$ (since $\mathtt{refl}_{\mathtt{a}}$ is identical to itself), and so by path induction it holds for all $\langle \mathtt{b}, \mathtt{p} \rangle$. Hence in an extensional theory we could use path induction to prove that all identifications are trivial. If the latter were so then the ∞-groupoid structure of higher identity types would be

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

rendered trivial as well, and so the connection with homotopy theory would be lost. Hence, the intensionality of HoTT — the distinction between external and internal identity — is essential to the homotopy interpretation.

3.4. Is "Identity" in HoTT Really Identity?

The radical innovations in the way identity is treated in HoTT make it worth questioning whether "identity" is the right name for the type $\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$.[14]

Maybe, what is called "internal identity" in HoTT is really some other kind of relation. Indeed it is often also called "equality", and it is notable that there is another relation called "identity", namely external identity. Furthermore much of the work on HoTT concerns what follows when a further principle about identity called the "Univalence Axiom" (UA) is added to the theory which relates identity and a weaker relation called "equivalence". In the next section, we briefly explain UA and consider its significance for the way identity is understood in HoTT. In particular, we relate UA to indiscernibility. We return to the question of this subsection in Section 5.

4. IDENTITY, EQUIVALENCE, AND UNIVALENCE

Up until now we have considered the identity relation between tokens of a given type. However, we also need to consider identity between types themselves. The identity relation between types is not something extra that needs to be added over and above that between tokens as described above, because types may themselves be regarded as tokens of a higher-order type or "universe". (For more on universes in HoTT, see [Ladyman and Presnell, forthcoming].) In other words, any two types $\mathtt{A}$ and $\mathtt{B}$ are tokens of some universe $\mathcal{U}$ (*i.e.*, $\mathtt{A} : \mathcal{U}$ and $\mathtt{B} : \mathcal{U}$), and thus we may form the identity type $\mathtt{Id}_{\mathcal{U}}(\mathtt{A},\mathtt{B})$ as we would for any tokens of any type. These identifications between types obey the same rules and behave exactly like identifications between tokens as described above.

One important difference between types and tokens in this respect is that we can use the resources of the theory to define other relations between arbitrary types that we cannot define between arbitrary tokens, and then consider how these relations interact with and relate to identity. A relation between types that is of particular importance is called "equivalence", which we define in the following subsection.

4.1. Equivalence

Equivalence in HoTT is a relation between types that is similar to the familiar relation of isomorphism. An isomorphism between types $\mathtt{A}$ and $\mathtt{B}$ (which in the HoTT Book is called a "quasi-inverse" [HoTT Book, Definition 2.4.6]) is

[14] Indeed we often encounter this reaction when talking about HoTT to people well-versed in logic and foundations of mathematics.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

a function $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ having a function $\mathtt{g} : \mathtt{B} \to \mathtt{A}$ that is both a pre- and post-inverse:[15]

$$(\mathtt{f} \circ \mathtt{g})(\mathtt{b}) = \mathtt{b} \qquad \text{for any } \mathtt{b} : \mathtt{B}$$
$$(\mathtt{g} \circ \mathtt{f})(\mathtt{a}) = \mathtt{a} \qquad \text{for any } \mathtt{a} : \mathtt{A}$$

(where we write $\mathtt{x} = \mathtt{y}$ to denote that the type $\mathtt{Id}_{\mathtt{A}}(\mathtt{x}, \mathtt{y})$ is inhabited, as noted in Section 2).

The relation called "equivalence" differs from isomorphism in a small but important way. Rather than requiring that a single function $\mathtt{g}$ serves as both pre- and post-inverse, $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ is an equivalence iff there exist $\mathtt{h}_1, \mathtt{h}_2 : \mathtt{B} \to \mathtt{A}$ such that $\mathtt{h}_1$ is a pre-inverse of $\mathtt{f}$ and $\mathtt{h}_2$ is a post-inverse of $\mathtt{f}$:

$$(\mathtt{f} \circ \mathtt{h}_1)(\mathtt{b}) = \mathtt{b} \qquad \text{for any } \mathtt{b} : \mathtt{B}$$
$$(\mathtt{h}_2 \circ \mathtt{f})(\mathtt{a}) = \mathtt{a} \qquad \text{for any } \mathtt{a} : \mathtt{A}.$$

It is not required that the two functions $\mathtt{h}_1$ and $\mathtt{h}_2$ be distinct, and so any function that is an isomorphism is also an equivalence. However, it is straightforward to prove [HoTT Book, Section 2.4] that if $\mathtt{f}$ has a pre-inverse $\mathtt{h}_1$ and a post-inverse $\mathtt{h}_2$ then each of these functions is also a post-inverse and a pre-inverse, certifying that $\mathtt{f}$ is an isomorphism.

So the difference between isomorphism and equivalence is not a matter of which functions satisfy the property. Rather the distinction is a "proof relevant" one, a matter of the evidence required to show that a function satisfies each property. The above two definitions can be formalised to give types $\mathtt{isIso}(\mathtt{f})$ and $\mathtt{isEquiv}(\mathtt{f})$ that say of a given function $\mathtt{f}$ that it is an isomorphism and an equivalence respectively. The principal difference between them is that it is possible to be an isomorphism in more than one way: if the type $\mathtt{isIso}(\mathtt{f})$ is inhabited, it may have multiple non-identical tokens. However, a function is an equivalence in essentially one way: if $\mathtt{f}$ is an equivalence then all tokens of $\mathtt{isEquiv}(\mathtt{f})$ are identical to one another.[16] (This is discussed in greater detail in [Ladyman and Presnell, forthcoming].)

Types $\mathtt{A}$ and $\mathtt{B}$ are **equivalent** iff there exists a function $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ satisfying $\mathtt{isEquiv}(\mathtt{f})$. Any such function is a token of the corresponding type $\mathtt{Equiv}(\mathtt{A}, \mathtt{B})$.[17]

4.2. Univalence

A major innovation of the HoTT research programme is the Univalence Axiom (UA) due to Vladimir Voevodsky, which relates (internal) identity between

[15] Pre- and post-inverse are more commonly called right- and left-inverses.

[16] We do not prove this here. The property of $\mathtt{f}$ defined above is called "bi-invertibility". A number of alternative formulations of the definition of equivalence are surveyed in [HoTT Book, Chapter 4].

[17] Strictly speaking, a token of $\mathtt{Equiv}(\mathtt{A}, \mathtt{B})$ is the function $\mathtt{f} : \mathtt{A} \to \mathtt{B}$ along with a token of $\mathtt{isEquiv}(\mathtt{f})$ demonstrating that it is an equivalence. But since, as noted above, all tokens of $\mathtt{isEquiv}(\mathtt{f})$ are identical, which particular token we have provides no further information and so may be suppressed.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

types to equivalence. It is taken to be one of the most important new discoveries arising from this approach to mathematics, and it is often said, for example by Steve Awodey, that UA gives rise to a "different understanding of identity":

> Rather than viewing it as identifying equivalent objects, and thus collapsing distinct objects, it is more useful to regard it as expanding the notion of identity to that of equivalence. [2014, p. 9]

As we might expect, given any identification between two types it is straightforward to show that those types are equivalent [HoTT Book, Lemma 2.10.1]. Thus we can define a function

$$\texttt{id-to-equiv}_{\texttt{A,B}} : \texttt{Id}_{\mathcal{U}}(\texttt{A},\texttt{B}) \to \texttt{Equiv}(\texttt{A},\texttt{B}).$$

However, from the basic rules of HoTT it is not in general possible to define a function in the opposite direction: types may be equivalent without also being provably identical.[18] The Univalence Axiom adds such a function

$$\texttt{equiv-to-id}_{\texttt{A,B}} : \texttt{Equiv}(\texttt{A},\texttt{B}) \to \texttt{Id}_{\mathcal{U}}(\texttt{A},\texttt{B})$$

taking equivalences between `A` and `B` as input, and returning identifications of `A` and `B` as output. Thus Univalence says that any two types that are equivalent may be identified.

In fact the Univalence Axiom says something stronger than this: it asserts that the function `id-to-equiv` is an equivalence, and so for any `A` and `B` the type $\texttt{Id}_{\mathcal{U}}(\texttt{A},\texttt{B})$ is equivalent to the type $\texttt{Equiv}(\texttt{A},\texttt{B})$, *i.e.*,

$$\texttt{Equiv}\big(\texttt{Id}_{\mathcal{U}}(\texttt{A},\texttt{B}),\ \texttt{Equiv}(\texttt{A},\texttt{B})\big).$$

In slogan form, Univalence is often expressed as 'Identity is equivalent to equivalence'. By applying `equiv-to-id` to the equivalence posited by UA we get an identity between $\texttt{Id}_{\mathcal{U}}(\texttt{A},\texttt{B})$ and $\texttt{Equiv}(\texttt{A},\texttt{B})$; so we could just as well express Univalence as 'Identity is identical to equivalence'.[19]

It is important to note that the identity relation involved throughout the discussion of Univalence is *internal* identity, not external identity. As noted above these are separate and distinct notions, and remain so under Univalence

[18] However, note that it is not possible to exhibit two types that are equivalent but provably non-identical. If this were possible then it would be inconsistent to add the Univalence Axiom to the system.

[19] Note that the full formal statement of Univalence involves a quantification over all pairs of types $\texttt{A},\texttt{B} : \mathcal{U}$, to say that the above equivalence holds for all types in universe $\mathcal{U}$ (*i.e.*, that universe $\mathcal{U}$ is "univalent"). However, the intended reading of this statement is a "typically ambiguous" one in which "$\mathcal{U}$" is taken as a dummy variable ranging over *all* universes. That is, the Univalence Axiom is not intended to assert that some particular universe is univalent, but rather that all universes are univalent. For more on this see [Shulman, 2012].

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

(or else the addition of UA would collapse the ∞-groupoid structure of identities, as discussed in Section 3.2). There are two points to note in this regard. First, recall that external identity entails internal identity, which in turn implies equivalence (even before UA is added). However, the addition of UA only gets us from equivalence to internal identity: we are not licensed to conclude that two equivalent types are *externally* identical. Second, the slogan noted above, 'Identity is identical to equivalence' should be understood as 'Internal identity is internally identical to equivalence'. Even after the addition of the Univalence Axiom, the notions of internal identity and equivalence remain distinct and are not externally identified.[20]

Univalence allows us to identify types that have essentially the same content presented in different ways. For example, without Univalence the product types $\mathtt{A} \times \mathtt{B}$ and $\mathtt{B} \times \mathtt{A}$ corresponding to the conjunctions $A \,\&\, B$ and $B \,\&\, A$ are not in general identical, whereas under Univalence the obvious equivalence between them gives an identification. For another example, under the basic rules of HoTT two functions $\mathtt{f}_1, \mathtt{f}_2 : \mathtt{A} \to \mathtt{B}$ may have the same input-output behaviour and yet not be provably identical. With the addition of Univalence it becomes possible to derive *function extensionality*, which says that functions are identical iff they have the same input-output behaviour (see [Gambino *et al.*, 2011] for a reconstruction of Voevodsky's derivation of this). So we might think of Univalence as restoring extensionality. However, as the above discussion makes clear, the degree of extensionality introduced by Univalence does not conflict with the essential intensionality of HoTT as embodied in the distinction between internal and external identity.

It might appear that the equivalence (and therefore identity) between the identity and equivalence relations means that one or other of these relations can be dispensed with. Perhaps UA means that identity does not need to be introduced as a primitive of the language, since it can be replaced with equivalence. This is not the case. Recall that the equivalence relation is only defined between types; there is no corresponding relation that can be defined between tokens of arbitrary types. Thus UA says only that identity and equivalence are equivalent *where both are defined*. It does not allow us to eliminate identity from the language, nor to reformulate HoTT in a manner that builds in Univalence by defining equivalence and having it play the role of identity (in standard HoTT at least — see footnote 20).

4.3. Justifying UA

Given that equivalence and identity remain externally distinct (even under the addition of Univalence), why should we take them to be internally identical by adopting the Univalence Axiom? After all, as Awodey acknowledges, Univalence

[20] This is the case in standard HoTT. In various versions of "cubical type theory" Univalence is built in so that equivalence and internal identity are externally identical [Licata and Brunerie, 2015].

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

'seems quite radical from a conventional foundational point of view, and it is indeed incompatible with classical foundations in set theory' [2014, p. 9].

One way we might justify Univalence is simply in terms of its mathematical effectiveness, by pointing to the useful results that follow as a consequence (such as the proof of function extensionality mentioned in Section 4.2). But while this may be an adequate answer from a mathematical point of view, it would be more satisfactory to have a more conceptual justification.

Awodey [2014] argues that UA is equivalent to the Principle of Invariance (PI) which says that all reasoning internal to the theory should be invariant under isomorphism. He relates Univalence to mathematical structuralism: if isomorphism is just sameness of structure, and Univalence says that isomorphic entities are identical, then it effectively says that 'mathematical objects simply *are* structures' [2014, p. 11]. However, this argument for Univalence is problematic. Recall from Section 4.2 that UA is expressed in terms of *equivalence*, which is subtly distinct from *isomorphism*. It turns out that the corresponding principle expressed in terms of isomorphism, *i.e.*,

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

is *inconsistent* [HoTT Book, Exercise 4.6]. The small difference between equivalence and isomorphism therefore makes a huge difference to the formulation of the Univalence Axiom, turning it from a powerful new principle into an inconsistency. In [Ladyman and Presnell, forthcoming] we examine this in more detail and explore arguments for taking equivalence rather than isomorphism to characterise "sameness of structure".

In the next section we consider a novel way to think about identity, under which Univalence is more easily justified.

5. "IDENTITY" AS INDISCERNIBILITY

As we have seen, the internal identity relation in HoTT represented by the identity types $\mathtt{Id}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ behaves radically differently from our intuitive view of identity in several ways:

- The very idea that there are *two* notions of identity in HoTT is already unusual. If external identity is the relation that says that two expressions "name the same token or type" and can be freely inter-substituted, how can some other relation also be "identity"? And how can two entities be distinct with respect to the former while "identical" with respect to the latter?
- Whereas conventionally identity is a binary yes/no property, in HoTT two tokens may have *multiple identifications*, some of which themselves may be identified, and so on.
- While internal identity is reflexive (and thus every token has a trivial self-identification) tokens may also have *non-trivial self-identifications*.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

- The *Univalence Axiom* says that two types may be identified in virtue of being isomorphic.

While these facts may be given mathematical justification and thus be rendered quite acceptable to the mathematician, they are nonetheless philosophically very puzzling and call into question the name "identity" for the relation (as noted in Section 3.4).

There is a natural alternative explanation that should be considered: that the relation represented by $\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$ is not identity at all but rather *indiscernibility*. In this section we outline what this position might mean in HoTT, before proceeding with a more detailed analysis in the next two sections.

If we were to understand $\mathtt{Id}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ as corresponding to indiscernibility instead of identity then the above features would no longer be unexpected, and would instead fit perfectly with our intuitions:

- If "internal identity" is understood as indiscernibility then external identity becomes the single "true" identity relation in the theory, and it is at least conceivable that two non-identical tokens might be indiscernible.
- The fact that two tokens are indiscernible may be witnessed in multiple different ways — for example, by different isomorphisms between them — and it is quite natural to study the correspondences between distinct isomorphisms.
- While any entity is trivially indiscernible from itself, if it has any non-trivial automorphisms then these provide additional ways in which the self-indiscernibility may be witnessed.
- On this view of $\mathtt{Id}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$, the Univalence Axiom says that types that are equivalent to each other (in a certain sense) are indiscernible from each other, and moreover that all witnesses to indiscernibility between types arise from such equivalences.[21]

Thus the idea that $\mathtt{Id}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ should be understood as representing indiscernibility is very appealing.

Aside from resolving the unusual characteristics of identity types, this "Identity-as-Indiscernibility" interpretation arguably has independent philosophical merits. Many philosophers have been bothered by the fact that we do not have direct access to facts about the identities of material objects. Rather, all we can obtain by empirical means are qualitative facts about the instantiation of properties by objects; so the best we can say is that two entities possess all the same qualitative properties. The obvious way to satisfy 'the empiricist sentiment that we ought to be able to have empirical access to individuation' [Ladyman *et al.*, 2012, p. 164] is to adopt the Principle of the Identity of Indiscernibles (PII), which says that any two entities that are alike in all their

[21] On this reading the Univalence Axiom can therefore more clearly be understood as an expression (or a strengthening) of the Principle of Invariance, as Awodey suggests [2014].

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

qualitative properties are identical. This principle reduces questions of identity to questions about empirically determinable facts. It may be understood as a kind of minimalism about identity: facts about individuation are exhausted by facts about indiscernibility.[22] (We examine PII in Section 6.6 and the Appendix.)

Returning to the domain of mathematics, we may apply corresponding reasoning when analysing the relation of internal identity. Whereas external identity corresponds to the idea of two entities being identical *by definition*, perhaps internal identity may be understood as the relation of *having all properties in common* and thus of being *indiscernible*. Naturally this relation is restricted to entities of the same type, since all properties that can be expressed pertain to entities of specific types. Then it seems natural to say, for example, that two topological spaces are identical (in this sense) if and only if they have all topological properties in common — even if they are defined differently.

Moreover it is plausible that, on some variants of Identity-as-Indiscernibility, axioms such as function extensionality and Univalence are not only natural but may follow as theorems. For example, rather than merely positing that two functions with the same input-output behaviour are equal, we may be able to prove the indiscernibility of any two functions whose input-output behaviour is indiscernible.

Having motivated the idea that "identity" might be understood as indiscernibility, in the remainder of the paper we consider three ways to implement it.

1. The first is as a new *interpretation* of HoTT, understanding $\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$ as indiscernibility and using this interpretation to justify the treatment of that type in HoTT (and in particular the Univalence Axiom, as discussed above).
2. The second is to explore the extent to which indiscernibility, formally defined, can play the role of identity in a variant of HoTT without identity types (the system we called "HoTT$^-$" in [Ladyman and Presnell, 2015]).
3. The third is to revise the definition of identity in HoTT to say that two tokens are identical iff they are indiscernible (*i.e.*, to adopt a version of PII).

In the following section we study the complications of defining a relation of indiscernibility within the language of HoTT — in particular those arising from the constructive logic. We then explore its relation to `Id`, in the form of the Principle of the Identity of Indiscernibles (PII) (Section 6.6 and Appendix). In Section 9 we consider the three approaches outlined above and find that all of them run into serious problems. In light of this we conclude that the "Identity-as-Indiscernibility" interpretation is not viable, and that identity, despite its

[22] Compare with the analogous position with regard to truth, that facts about truth are exhausted by the T-schema.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

unusual features, must be taken as primitive, as is standardly supposed in work on HoTT.

6. DISCERNIBILITY AND INDISCERNIBILITY

In order to study discernibility and indiscernibility in HoTT we must formally define them. The underlying idea that we are trying to capture is the one outlined in Section 5, namely the possibility of discriminating between tokens using properties expressible in the language. Like identity, these relations are therefore restricted to pairs of tokens of the same type, since tokens of distinct types do not share a common vocabulary of properties to be compared.

The basic ideas of discernibility and indiscernibility are easy to express in the language of HoTT. However, since the logic of HoTT is constructive the relationship between them is not as straightforward as it would be in classical logic. While the two notions are clearly contrary to one another (*i.e.*, no two tokens can be both discernible and indiscernible) it is not clear whether they should be given separate definitions or if one should be defined as the negation of the other. We investigate these questions below.

6.1. Indiscernibility and Language

Since the idea of (in)discernibility involves distinguishing tokens by means of properties expressible in the language, we must pay attention to what language we are working in. If the language includes identity types, and the properties quantified over include *all* properties definable in the language, then indiscernibility reduces to identity. For any token $\mathtt{a} : \mathtt{A}$ we can define the "is identical to $\mathtt{a}$" predicate $\mathtt{Id}_{\mathtt{A}}(\mathtt{a})$ that maps any token $\mathtt{x} : \mathtt{A}$ to the type $\mathtt{Id}_{\mathtt{A}}(\mathtt{a}, \mathtt{x})$. This predicate is obviously satisfied by $\mathtt{a}$ itself. Thus for any token $\mathtt{b}$ that satisfies all the same predicates as $\mathtt{a}$ we have $\mathtt{Id}_{\mathtt{A}}(\mathtt{a}, \mathtt{b})$.[23]

We might try to define the notion of *non-identity-involving* (NII) predicates within HoTT. The obvious way to (try to) do this would be to define NII recursively, with some base cases and some rules for construction. If we could do this correctly then perhaps we could restrict the quantification only to NII predicates.

Alternatively, we could restrict attention to the fragment of HoTT without identity types (which we called HoTT$^-$ in [Ladyman and Presnell, 2015]) and consider the definition of (in)discernibility in this language. However, this would mean that we could not formally compare identity and indiscernibility to see, for example, whether the Principle of Identity of Indiscernibles (PII) holds in HoTT.

However, both these strategies suffer from the same problem. In HoTT (and HoTT$^-$) we are free to supplement the basic language with any further types we wish to consider. For example, the existence of the natural numbers is not part of the basic language but we may define a type $\mathbb{N}$ with appropriate token constructors and elimination rule. Thus even though $\mathtt{Id}_{\mathtt{A}}(\mathtt{x}, \mathtt{y})$ is not part of the

[23] *I.e.*, the identity predicate provides **haecceities** for all tokens.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

basic language of HoTT$^-$, nothing prevents us from supplementing it with a defined type that has exactly the same behaviour as the identity type.

Similarly, however we define what it means for a predicate to be "non-identity-involving", it seems impossible to eliminate the possibility that a predicate may "sneak identity back in" by reference to additional types defined within the language that (directly or cryptically) correspond to identity types. The only way to rule this out completely, it seems, is to disallow NII predicates from making reference to any defined types beyond those constructed from the basic language itself. But this would make the notion of "NII predicate" (and the corresponding relations of discernibility and indiscernibility) far too weak, since it would prevent us from discerning tokens by means of, for example, functions that map into the natural numbers.

The conclusion, then, is that strictly speaking indiscernibility in the sense in which we are interested, namely in which the quantification is over predicates that do not involve identity either directly nor cryptically, *cannot* be formally defined in HoTT or HoTT$^-$. However, rather than immediately abandoning the idea of Identity-as-Indiscernibility we instead proceed "semi-formally". Whenever we quantify over predicates we intend this to be understood as ranging over only the non-identity-involving predicates (even though this restriction cannot be formally defined in the language). Similarly, any constructions involved in proofs or counterexamples should not depend upon the use of identity-involving predicates (except in case where the use of the identity is itself the point of the proof or counterexample).

With this caveat in mind we now proceed to consider how discernibility and indiscernibility should be defined.

6.2. What Kind of Discernibility?

The philosophical literature on discernibility (going back to at least [Quine, 1976], and more recently in, for example, [Ladyman *et al.*, 2012; Caulton and Butterfield, 2012]) has considered a variety of ways in which entities may be said to be distinguished. In particular, two entities are

- '**absolutely discernible** when there is a property that one object has that the other lacks.'
- '**weakly discernible** when there is a relation in which the first stands to the second but the first does not stand to itself.' [Ladyman *et al.*, 2012, pp. 164–165] (emphasis added)

In HoTT a property that is assertible of things of type $\mathtt{A}$ is defined by a function $\mathtt{P} : \mathtt{A} \to \mathcal{U}$, which maps each token $\mathtt{x} : \mathtt{A}$ to the type $\mathtt{P}(\mathtt{x})$ corresponding to the proposition that this token satisfies the property. Similarly, a binary relation $\mathtt{R}$ maps a pair of tokens $\mathtt{x}, \mathtt{y}$ to the type $\mathtt{R}(\mathtt{x}, \mathtt{y})$ asserting that this pair stands in the relation.

Since in HoTT all multi-input functions are defined by currying (as in lambda calculus), any binary relation $[\mathtt{x}, \mathtt{y} \mapsto \mathtt{R}(\mathtt{x}, \mathtt{y})]$ may be partially applied to a single

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

input $\mathtt{a} : \mathtt{A}$ to give a one-place predicate $[\mathtt{y} \mapsto \mathtt{R}(\mathtt{a}, \mathtt{y})]$ (and more generally any n-place predicate may be partially applied to some of its inputs to give an n'-place predicate for $n' < n$). Thus there is no sharp distinction between two-place relations and one-place predicates.[24]

In particular this means that *weak discernibility* collapses to *absolute discernibility*: if there is a relation $\mathtt{W}$ such that $\mathtt{W}(\mathtt{x}, \mathtt{y})$ holds but $\mathtt{W}(\mathtt{x}, \mathtt{x})$ does not, then by partial application to $\mathtt{x}$ we obtain a one-place property $\mathtt{W}_\mathtt{x}$ such that $\mathtt{W}_\mathtt{x}(\mathtt{y})$ holds but $\mathtt{W}_\mathtt{x}(\mathtt{x})$ does not.

There are many other notions of discernibility that may be defined, such as "relative discernibility" which holds 'when there is a two-place relation in which the first stands to the second but the second does not stand to the first'. However, it is proved in [Ladyman *et al.*, 2012] that weak discernibility is 'the most discerning nontrivial discernibility relation' (p. 184), meaning that if two elements may be discerned by one of these relations then they are also weakly discernible. Thus, since in HoTT weak discernibility collapses to absolute discernibility, the latter is the natural definition of discernibility to use and there is no need to consider the other relations.

6.3. Discernibility

We may formalise *absolute discernibility* as:

$$\mathtt{Dis}_\mathtt{A}(\mathtt{a}, \mathtt{b}) :\equiv \sum_{\mathtt{P}:\mathtt{A}\to\mathcal{U}} \mathtt{P}(\mathtt{a}) \times \neg\mathtt{P}(\mathtt{b})$$

which asserts that there is some property $\mathtt{P}$ that holds of $\mathtt{a}$ and not of $\mathtt{b}$. Tokens of this type are triples $\langle \mathtt{P}, \mathtt{p}_\mathtt{a}, \bar{\mathtt{p}}_\mathtt{b} \rangle$, where $\mathtt{P} : \mathtt{A} \to \mathcal{U}$ is a predicate on $\mathtt{A}$, token $\mathtt{p}_\mathtt{a} : \mathtt{P}(\mathtt{a})$ demonstrates that $\mathtt{a}$ satisfies this predicate, and $\bar{\mathtt{p}}_\mathtt{b} : \neg\mathtt{P}(\mathtt{b})$ demonstrates that $\mathtt{b}$ does not.

This relation is obviously irreflexive: a triple $\langle \mathtt{P}, \mathtt{p}_\mathtt{a}, \bar{\mathtt{p}}_\mathtt{a} \rangle : \mathtt{Dis}_\mathtt{A}(\mathtt{a}, \mathtt{a})$ leads immediately to contradiction. Thus we can trivially prove

$$\prod_{\mathtt{a}:\mathtt{A}} \neg\mathtt{Dis}_\mathtt{A}(\mathtt{a}, \mathtt{a})$$

as expected. Also, note that despite the asymmetry in the statement of the definition this relation is symmetric, since if we have $\langle \mathtt{Q}, \mathtt{q}_\mathtt{a}, \bar{\mathtt{q}}_\mathtt{b} \rangle : \mathtt{Dis}_\mathtt{A}(\mathtt{a}, \mathtt{b})$ then we also have $\langle \neg\mathtt{Q}, \bar{\mathtt{q}}_\mathtt{b}, \bar{\bar{\mathtt{q}}}_\mathtt{a} \rangle : \mathtt{Dis}_\mathtt{A}(\mathtt{b}, \mathtt{a})$, where $\bar{\bar{\mathtt{q}}}_\mathtt{a} : \neg\neg\mathtt{Q}(\mathtt{a})$ is derived from $\mathtt{q}_\mathtt{a} : \mathtt{Q}(\mathtt{a})$ by Double-Negation Introduction.

6.4. Indiscernibility

An obvious way to define indiscernibility is simply as the negation of discernibility, $\neg\mathtt{Dis}_\mathtt{A}(\mathtt{a}, \mathtt{b})$. However, since the logic we are working in is constructive

[24] Note that in general a predicate being one-placed says nothing about the metaphysics of the corresponding property. For example, the property of being an uncle is relational, not intrinsic, but corresponds to the one-place predicate "is an uncle".

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

the introduction of negations into definitions tends to weaken them, and so we should investigate whether a stronger definition can be given. (We investigate the relationship between indiscernibility and discernibility in constructive logic in Section 6.5.)

Whereas a negative characterisation of indiscernibility says that there is no property that discriminates the two tokens (which is what $\neg\mathtt{Dis}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ says), a positive characterisation instead says that all properties are shared by the two tokens. In other words, if one token has some property then so does the other. One way to express this would be:

$$\mathtt{Indis}_{\mathtt{A}}(\mathtt{a},\mathtt{b}) :\equiv \prod_{\mathtt{P}:\mathtt{A}\to\mathcal{U}} \mathtt{P}(\mathtt{a}) \leftrightarrow \mathtt{P}(\mathtt{b}),$$

where the notation "$\mathtt{X} \leftrightarrow \mathtt{Y}$" is an abbreviation for $(\mathtt{X} \to \mathtt{Y}) \times (\mathtt{Y} \to \mathtt{X})$. (We consider an alternative statement of indiscernibility below.)

Clearly $\mathtt{Indis}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ is symmetric, reflexive, and transitive. Furthermore, it supports substitution: if we have $\mathtt{Indis}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ and we have $\mathtt{Q}(\mathtt{a})$ for some property $\mathtt{Q}$ then we can derive $\mathtt{Q}(\mathtt{b})$. It is therefore of interest to see what parallels can be drawn between the behaviour of $\mathtt{Id}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ and $\mathtt{Indis}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$, which we do in Sections 6.6 and 9.

One might object that that above definition of indiscernibility is too weak, because $\leftrightarrow$ is not the appropriate relationship to demand between $\mathtt{P}(\mathtt{a})$ and $\mathtt{P}(\mathtt{b})$. Since in HoTT we work proof-relevantly, we care not just that both $\mathtt{a}$ and $\mathtt{b}$ satisfy the property but also about the structure of the respective types. For example, consider a property $\mathtt{Y}$ such that $\mathtt{a}$ satisfies the property in exactly one way (*i.e.*, $\mathtt{Y}(\mathtt{a}) \simeq \mathtt{1}$) while $\mathtt{b}$ satisfies it in a countable infinity of ways (*i.e.*, $\mathtt{Y}(\mathtt{b}) \simeq \mathbb{N}$). Then, although we have $\mathtt{Y}(\mathtt{a}) \leftrightarrow \mathtt{Y}(\mathtt{b})$ since both types are inhabited, it seems incorrect to say that $\mathtt{a}$ and $\mathtt{b}$ are not discerned by $\mathtt{Y}$ in some way, since by inspecting $\mathtt{Y}(\mathtt{a})$ and $\mathtt{Y}(\mathtt{b})$ we could tell which was which. Thus we might instead consider a different notion of indiscernibility:

$$\mathtt{Indis}'_{\mathtt{A}}(\mathtt{a},\mathtt{b}) :\equiv \prod_{\mathtt{P}:\mathtt{A}\to\mathcal{U}} \mathtt{Equiv}(\mathtt{P}(\mathtt{a}),\mathtt{P}(\mathtt{b})).$$

Clearly, $\mathtt{Indis}'_{\mathtt{A}}(\mathtt{a},\mathtt{b}) \to \mathtt{Indis}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$, since an equivalence between $\mathtt{P}(\mathtt{a})$ and $\mathtt{P}(\mathtt{b})$ just consists in a pair of functions $\mathtt{P}(\mathtt{a}) \to \mathtt{P}(\mathtt{b})$ and $\mathtt{P}(\mathtt{b}) \to \mathtt{P}(\mathtt{a})$ satisfying some conditions. We can prove that $\mathtt{Indis}_{\mathtt{A}}(\mathtt{a},\mathtt{b}) \to \mathtt{Indis}'_{\mathtt{A}}(\mathtt{a},\mathtt{b})$. Given any property $\mathtt{P} : \mathtt{A} \to \mathcal{U}$ and any $\mathtt{a} : \mathtt{A}$ we can define a property $\mathtt{Q}_{\mathtt{P},\mathtt{a}} : \mathtt{A} \to \mathcal{U}$ by

$$\mathtt{Q}_{\mathtt{P},\mathtt{a}}(\mathtt{x}) :\equiv \mathtt{Equiv}(\mathtt{P}(\mathtt{a}),\mathtt{P}(\mathtt{x})).$$

$\mathtt{Q}_{\mathtt{P},\mathtt{a}}(\mathtt{a})$ holds trivially, and from $\mathtt{Indis}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ we have $\mathtt{Q}_{\mathtt{P},\mathtt{a}}(\mathtt{a}) \leftrightarrow \mathtt{Q}_{\mathtt{P},\mathtt{a}}(\mathtt{b})$; so we have $\mathtt{Equiv}(\mathtt{P}(\mathtt{a}),\mathtt{P}(\mathtt{b}))$ as required.[25]

[25] Note that this construction clarifies the sense in which $\mathtt{a}$ and $\mathtt{b}$ are discerned by $\mathtt{Y}$ in the example above.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

Thus although $\mathtt{Indis}'_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ is "more natural" from a HoTT perspective, the two definitions are of equal strength, and so we use the simpler definition.

6.5. Indiscernibility and Discernibility

In this subsection we consider the relationship between $\mathtt{Dis}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ and $\mathtt{Indis}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ (which we abbreviate as "Dis" and "Indis" for the present discussion). First, it is clear that they are contrary to one another, $(\mathtt{Indis} \times \mathtt{Dis}) \to \mathtt{0}$, and so each entails the negation of the other:

$$\mathtt{Indis} \vdash \neg\mathtt{Dis} \quad \text{and} \quad \mathtt{Dis} \vdash \neg\mathtt{Indis}.$$

Furthermore, for any $\mathtt{x} : \mathtt{A}$ we can define the predicate $\mathtt{Q}_{\mathtt{x}}$, where

$$\mathtt{Q}_{\mathtt{x}}(\mathtt{y}) :\equiv \mathtt{Indis}_{\mathtt{A}}(\mathtt{x},\mathtt{y}).$$

For any $\mathtt{a}$ and $\mathtt{b}$ such that $\neg\mathtt{Indis}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$, we have $\mathtt{Q}_{\mathtt{a}}(\mathtt{a})$ and $\neg\mathtt{Q}_{\mathtt{a}}(\mathtt{b})$, and so $\mathtt{Q}_{\mathtt{a}}$ itself discerns $\mathtt{a}$ from $\mathtt{b}$. Thus $\neg\mathtt{Indis} \vdash \mathtt{Dis}$, and so we have

$$\neg\mathtt{Indis} \dashv\vdash \mathtt{Dis}$$

It is then straightforward to show that

$$\neg\mathtt{Indis} \dashv\vdash \neg\neg\mathtt{Dis}$$
$$\neg\neg\mathtt{Indis} \dashv\vdash \neg\mathtt{Dis}.$$

So, in particular, we have $\neg\neg\mathtt{Dis} \vdash \mathtt{Dis}$.[26]

6.6. Identity and Indiscernibility

Classically, the relationship between identity and indiscernibility may be split into two components: the Indiscernibility of Identicals and the Principle of Identity of Indiscernibles (PII).

The former principle is easily dealt with. It is a straightforward consequence of path induction [HoTT Book, Section 1.12] that identical tokens share all their properties: for any type $\mathtt{A}$ and any predicate $\mathtt{P} : \mathtt{A} \to \mathcal{U}$, there is a function

$$\mathtt{f} : \prod_{\mathtt{x},\mathtt{y}:\mathtt{A}} \prod_{\iota:\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})} \mathtt{P}(\mathtt{x}) \leftrightarrow \mathtt{P}(\mathtt{y})$$

and in particular for any $\mathtt{a} : \mathtt{A}$

$$\mathtt{f}(\mathtt{a},\mathtt{a},\mathtt{refl}_{\mathtt{a}}) :\equiv (1_{\mathtt{P}(\mathtt{a})}, 1_{\mathtt{P}(\mathtt{a})})$$

[26] We are presently not able to answer whether Indis follows from ¬Dis. From ¬Dis we immediately derive $\prod_{\mathtt{P}:\mathtt{A}\to\mathcal{U}} \neg(\mathtt{P}(\mathtt{a}) \times \neg\mathtt{P}(\mathtt{b}))$ but since $\neg(\mathtt{X} \times \neg\mathtt{Y})$ is in general weaker than $(\mathtt{X} \to \mathtt{Y})$, it does not follow that we can derive Indis (although there may be some alternative derivation that we have not considered).

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

(where for any type $\mathtt{X}$ the function $1_{\mathtt{X}} : \mathtt{X} \to \mathtt{X}$, usually called the "identity function", leaves its input unchanged). Thus identical tokens are indiscernible

$$\mathtt{Id} \vdash \mathtt{Indis}$$

(and so of course external identity also implies indiscernibility). The contrapositive of the Indiscernibility of Identicals, $\neg\mathtt{Indis} \vdash \neg\mathtt{Id}$, is called the principle of the "Diversity of the Dissimilar" [McTaggart, 1921, Chapter 10].[27] Due to the mutual entailment between $\neg\mathtt{Indis}$ and $\mathtt{Dis}$ shown above it may be re-written as $\mathtt{Dis} \vdash \neg\mathtt{Id}$.

The study of PII is more complex. Its original formulation by Leibniz is: 'it is not true that two substances may be exactly alike and differ only numerically, solo numero' [1686/1962, Section 9], or in other words *there are no two (non-identical) objects that share all their properties.*[28]

The most direct translation of this is:

$$\neg \sum_{\mathtt{x},\mathtt{y}:\mathtt{A}} \big(\mathtt{Indis}_{\mathtt{A}}(\mathtt{x},\mathtt{y}) \times \neg\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})\big),$$

where $\mathtt{Indis}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$ expresses that $\mathtt{x}$ and $\mathtt{y}$ are 'exactly alike' (*i.e.*, have all properties in common), and we take 'differ only numerically, solo numero' to mean simply that $\mathtt{x}$ and $\mathtt{y}$ are not identical. We may also write this as

$$\prod_{\mathtt{x},\mathtt{y}:\mathtt{A}} \neg\big(\mathtt{Indis}_{\mathtt{A}}(\mathtt{x},\mathtt{y}) \times \neg\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})\big)$$

which is constructively equivalent to the above and says that 'any two tokens may not be indiscernible and yet non-identical'.

Classically we may go a step further and re-write this again as 'any two indiscernible things are in fact identical' (which is how PII is usually stated), which we may formalise as:

$$\prod_{\mathtt{x},\mathtt{y}:\mathtt{A}} \big(\mathtt{Indis}_{\mathtt{A}}(\mathtt{x},\mathtt{y}) \to \mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})\big).$$

But constructively this is a stronger claim, since for arbitrary types $\mathtt{P}$ and $\mathtt{Q}$ we have $\mathtt{P} \to \mathtt{Q} \vdash \neg(\mathtt{P} \times \neg\mathtt{Q})$ but not the reverse entailment.

Hence, in principle PII could be formulated in different ways in HoTT, and we might wonder which formulation(s) we should expect to hold (although it is clear that we should not expect indiscernible tokens to be *externally* identical). Various possibilities are presented and briefly analysed in the Appendix and we conclude that the appropriate one is the stronger version immediately above, which we abbreviate as $\mathtt{Indis} \to \mathtt{Id}$.

[27] The move from $\mathtt{X} \vdash \mathtt{Y}$ to $\neg\mathtt{Y} \vdash \neg\mathtt{X}$ is constructively valid, as is easily shown.

[28] Leibniz gives an argument from haecceities in the *Discourse on Metaphysics*. He also argues for PII from the Principle of Sufficient Reason: if objects differed solo numero then God would have no reason to arrange them one way rather than another.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

Finally, we should also consider the contrapositive of the (usual statement of) PII, namely the principle of the 'Dissimilarity of the Diverse' [McTaggart, 1921, p. 101]. With PII formulated as $\mathtt{Indis} \to \mathtt{Id}$, its contrapositive is written as $\neg\mathtt{Id} \to \neg\mathtt{Indis}$, or (again using the mutual entailment between $\neg\mathtt{Indis}$ and $\mathtt{Dis}$) as

$$\neg\mathtt{Id} \to \mathtt{Dis},$$

i.e., any tokens that are provably non-identical may be discerned.[29]

The difficulties discussed in the Appendix as to the correct formulation of PII apply likewise to this principle.

In the next section we consider whether finite types violate PII in even the weakest form, as might seem plausible.

7. FINITE TYPES AND CHARACTERISTIC FUNCTIONS

The debate about mathematical structuralism (see, in particular, [Leitgeb and Ladyman, 2008] and [Shapiro, 2008]) seems to have concluded that PII, even in its formulation in terms of weak discernibility, fails to hold for some mathematical structures such as edgeless graphs, because arguably they contain elements that are utterly indiscernible. Specifically, then, they violate the principle of the Dissimilarity of the Diverse — they are numerically diverse but fail to be dissimilar with respect to any property or relation.

We might therefore expect a similar phenomenon to arise for finite types in HoTT, *i.e.*, types that have a finite number of token constructors all of which take no inputs (such as the Unit type $\mathtt{1}$). That is, we might suspect that finite types give an example of internally distinct tokens that are not discerned by any predicate, and thus violate the principle of the Dissimilarity of the Diverse. In this section we consider this and related issues and find that the finite types do not produce a counterexample to PII.

7.1. Finite Types

The basic language of HoTT provides the Unit type, which has a single token constructor producing token $* : \mathtt{1}$. Similarly, we may define the **binary type** $\mathtt{2}$ having two token constructors that produce tokens $\alpha : \mathtt{2}$ and $\beta : \mathtt{2}$. By definition, tokens arising from different constructors are externally distinct, and it is possible (but not straightforward, see [HoTT Book, Section 2.12]) to prove that they are internally distinct as well (*i.e.*, $\mathtt{Id}_{\mathtt{2}}(\alpha, \beta) \to \mathtt{0}$).[30]

We first consider the situation without assuming Univalence. To discern α and β we need a predicate of type $\mathtt{2} \to \mathcal{U}$ that holds of one and not the other.

[29] Note that, as throughout this discussion, the notion of identity we are concerned with here is *internal* identity. As we would expect in an intentional theory, the corresponding statement for external identity is of course routinely violated, since two tokens may be externally distinct but internally identified and therefore not discernible.

[30] Alternatively we could consider the coproduct type $\mathtt{1} + \mathtt{1}$ which also has two tokens that are externally and internally distinct.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

Recall that predicates on a given type, like all functions, are given by the recursion and induction rules for that type. In the present case, the recursion rule for $\mathtt{2}$ [HoTT Book, Section 1.8] says that to define a predicate $\mathtt{P} : \mathtt{2} \to \mathcal{U}$ we must pick out two types $\mathtt{P}(\alpha)$ and $\mathtt{P}(\beta)$, and any choice of two types gives a predicate on $\mathtt{2}$. In other words, a predicate on $\mathtt{2}$ is completely specified in terms of what it does with tokens α and β.

In particular, let $\mathtt{P}_\alpha$ be the predicate that maps α to the Unit type $\mathtt{1}$ and β to the Zero type $\mathtt{0}$, and vice versa for $\mathtt{P}_\beta$. Clearly we have a token of $\mathtt{P}_\alpha(\alpha) \times \neg\mathtt{P}_\alpha(\beta)$ and a token of $\mathtt{P}_\beta(\beta) \times \neg\mathtt{P}_\beta(\alpha)$. These "characteristic predicates" therefore discern α and β.

More generally, we may define finite types having any finite number of token constructors producing externally and internally distinct tokens $\mathtt{a}, \mathtt{b}, \mathtt{c}, \ldots$ The recursion rule for each finite type $\mathtt{F}$ then says that to define a predicate $\mathtt{Q} : \mathtt{F} \to \mathcal{U}$ we must pick out a type $\mathtt{Q}(\mathtt{a}), \mathtt{Q}(\mathtt{b}), \mathtt{Q}(\mathtt{c}), \ldots$ for each constructor, and that any such choice defines a predicate. Thus for any finite type $\mathtt{F}$ and for each constructor producing a token $\mathtt{t} : \mathtt{F}$ we can define a characteristic predicate $\mathtt{P}_\mathtt{t}$ that returns $\mathtt{1}$ when given $\mathtt{t}$ and returns $\mathtt{0}$ when given any other token of $\mathtt{F}$.

Thus by the use of characteristic predicates all externally distinct tokens of finite types can be discerned — for tokens $\mathtt{x}$, $\mathtt{y}$ of any finite type, if $\mathtt{x} \not\equiv \mathtt{y}$ then $\mathtt{Dis}(\mathtt{x}, \mathtt{y})$ (and of course, if $\mathtt{x} \equiv \mathtt{y}$ then $\mathtt{Indis}(\mathtt{x}, \mathtt{y})$, trivially). In this sense it follows from the rules for defining functions on finite types that the facts about the external identity and distinctness of tokens of those types are reflected into theorems about their indiscernibility and discernibility (respectively).

One might object that this discernment is somehow illicitly using identity because the characteristic predicates need to be able to tell the difference between distinct input tokens in order to give the right output when given a particular input. However, recall that predicates, like all functions, are defined via substitution on expressions. Thus predicates are not given as their inputs the tokens α and β but rather *expressions denoting* α and β. Hence, this kind of discernment is via *names*. In HoTT every token of every type has a name which is an expression that refers uniquely to that token and to no other. Of course, the name of a token is arbitrary, and could be changed to any other expression (as long as this is done consistently and without collision with any expression already in use). We are assuming that the distinctness and discernibility of expressions is not uncertain, since this is a basic requirement of being able to use expressions at all. Thus the definition of characteristic predicates does depend upon a pre-existing notion of identity, but this is the *external* identity relation $\equiv$ that holds between expressions, rather than the internal propositional identity represented by identity types.

In sum, in the absence of Univalence, even the strongest form of PII holds for all finite types in HoTT since all the elements of a finite type are distinguished by having different names, and one of the defining properties of a finite type is that functions can be defined by case analysis over those generating names.

If Univalence is assumed the situation is complicated by the fact that we may have externally distinct but internally identical tokens of finite types (since

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

they can be generated by permutations which, being equivalences, give new identifications under Univalence).[31] Hence in the presence of Univalence it is no longer true that facts about indiscernibility reflect facts about external identity. However, since these tokens are internally identical they do not generate a violation of PII.

7.2. Generalising from Finite Types

We turn now from finite types to types in general. For some types $\mathtt{X}$, predicates on $\mathtt{X}$ are defined via the recursion/induction rules, much as with finite types. In such cases we will be able to define characteristic predicates for each token. For example, for the natural numbers $\mathbb{N}$ (with zero element $0_{\mathbb{N}} : \mathbb{N}$ and successor function $\mathtt{s} : \mathbb{N} \to \mathbb{N}$) we can define the following family of predicates $\chi_{\mathtt{i}} : \mathbb{N} \to \mathcal{U}$ (for $\mathtt{i} : \mathbb{N}$)

$$\begin{aligned} \chi_{0_{\mathbb{N}}}(0_{\mathbb{N}}) &:\equiv \mathtt{1} \\ \chi_{0_{\mathbb{N}}}(\mathtt{s}(\mathtt{m})) &:\equiv \mathtt{0} \\ \chi_{\mathtt{s}(\mathtt{n})}(0_{\mathbb{N}}) &:\equiv \mathtt{0} \\ \chi_{\mathtt{s}(\mathtt{n})}(\mathtt{s}(\mathtt{m})) &:\equiv \chi_{\mathtt{n}}(\mathtt{m}) \end{aligned}$$

Thus for any $\mathtt{n} : \mathbb{N}$ we have $\chi_{\mathtt{n}}(\mathtt{n}) \equiv \mathtt{1}$, and $\chi_{\mathtt{n}}(\mathtt{j}) \equiv \mathtt{0}$ for all $\mathtt{j} \not\equiv \mathtt{n}$, so $\chi_{\mathtt{n}}$ is a characteristic predicate for $\mathtt{n}$.

The recursion rule for an arbitrary type $\mathtt{X}$ will (in most cases) say that in order to define a predicate (or any function) on $\mathtt{X}$ it suffices to specify what that predicate does to the outputs of the constructors for $\mathtt{X}$ [HoTT Book, Section 5.6]. Thus arbitrary *constructed tokens* (*i.e.*, tokens arising as the output of a constructor) can in general be discerned by characteristic predicates. However, we cannot assume that *every* token is externally identical to the output of a constructor, and so it does not follow that all tokens can be discerned, but we have not been able to produce a counterexample to PII.[32]

---

[31] Consider a permutation on a finite type, such as the $\mathtt{swap}$ function on the binary type $\mathtt{2}$ defined as $\mathtt{swap}(\alpha) :\equiv \beta$ and $\mathtt{swap}(\beta) :\equiv \alpha$. This is a non-trivial autoequivalence of the type, and so from Univalence we obtain a non-trivial self-identification of $\mathtt{2}$ with itself, $\mathtt{UA}(\mathtt{swap}) : \mathtt{Id}_{\mathcal{U}}(\mathtt{2}, \mathtt{2})$. Let $[\mathtt{X} \mapsto \mathtt{X}] : \mathcal{U} \to \mathcal{U}$ be the null operation that returns its input unchanged. The transport [HoTT Book, Lemma 2.3.1] of this operation along the identification $\mathtt{UA}(\mathtt{swap})$ produces a function $\mathtt{tr}^{[\mathtt{X} \to \mathtt{X}]}(\mathtt{UA}(\mathtt{swap})) : \mathtt{2} \to \mathtt{2}$. Thus applying this function to $\alpha : \mathtt{2}$ gives a token of $\mathtt{2}$. While this token can be shown to be internally identical to $\beta : \mathtt{2}$, it is not externally identical to it. We are grateful to an anonymous referee for pointing this out to us.

[32] The types discussed above, for which predicates may be defined via the recursion/induction rules, are called "positive" types. The function type $\mathtt{A} \to \mathtt{B}$ is an example of a "negative type" for which there is no recursion rule; instead the elimination rule tells us how to apply a function to an argument. For more on this distinction see [HoTT Book, Section 2.12]. We are grateful to an anonymous referee for this journal for drawing our attention to this point.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

8. CASE ANALYSIS

In this section we briefly consider the relationship between case analysis and indiscernibility.

Function (and predicate) definitions may be written in "case analysis" or "pattern matching" style; for example, the definition of the predicate $\mathtt{P}_\alpha : 2 \to \mathcal{U}$ considered in Section 7.1 may be written as

$$\mathtt{P}_\alpha(\alpha) :\equiv \mathtt{1}$$
$$\mathtt{P}_\alpha(\beta) :\equiv \mathtt{0}.$$

Such a definition might be understood as standing for an "if ... then ..." expression; specifically, either

$$\text{if } \mathtt{x} \equiv \alpha \text{ then } \mathtt{P}_\alpha(\mathtt{x}) :\equiv \mathtt{1}\text{; else if } \mathtt{x} \equiv \beta \text{ then } \mathtt{P}_\alpha(\mathtt{x}) :\equiv \mathtt{0}$$

or, if we have identity types in the language, perhaps as

$$\text{if } \mathtt{x} = \alpha \text{ then } \mathtt{P}_\alpha(\mathtt{x}) :\equiv \mathtt{1}\text{; else if } \mathtt{x} = \beta \text{ then } \mathtt{P}_\alpha(\mathtt{x}) :\equiv \mathtt{0}.$$

One might object to such a definition on the grounds that it only says what to do with the constructed tokens, and says nothing about how to handle arbitrary tokens. In order for such a definition to be adequate we need some assurance of *exhaustivity* — either that no other tokens of the type exist, or that the value of $\mathtt{P}_\alpha$ at any other token is fully determined by the above specification.

In HoTT this may be resolved by proving a "Uniqueness Principle" for the type, which (in many cases) says that every token of the type is equal to some constructed token.[33]

For example, for the finite type 2 described above the Uniqueness Principle says [HoTT Book, Equation 1.8.1]

$$\prod_{\mathtt{x}:2} \mathtt{Id}_2(\mathtt{x}, \alpha) + \mathtt{Id}_2(\mathtt{x}, \beta).$$

Thus given any arbitrary token $\mathtt{x}$ of the type, the Uniqueness Principle gives an identification between $\mathtt{x}$ and some constructed token $\mathtt{c}$, and then from this identification we derive an identification between $\mathtt{f}(\mathtt{x})$ and $\mathtt{f}(\mathtt{c})$, where $\mathtt{f}(\mathtt{c})$ is given explicitly by the function definition.[34] Thus the Uniqueness Principle ensures that although the function is only explicitly given values at constructed

[33] This is not always what the Uniqueness Principle for a type says: for example, for identity types the Uniqueness Principle does not say that every identification is equal to a trivial self-identification (which statement is not well-typed, as was noted in Section 3.3), but rather that every identification has a counterpart in a based identity type that is equal to the counterpart of a trivial self-identification. See [Ladyman and Presnell, 2015, Section 6.2].

[34] A similar situation holds when the Uniqueness Principle does not give an identification between arbitrary tokens and constructed tokens, as discussed in footnote 33.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

tokens, it is nonetheless well-defined for all tokens of the type, up to internal identity.

Since Uniqueness Principles directly involve identifications they of course cannot be stated in a language without identity types, and so this assurance of exhaustivity is not available. However, wherever in HoTT we could prove a Uniqueness Principle for a type saying that every token of the type is *equal* to a constructed token, we should be able to prove a corresponding "Indiscernibility Principle" saying that every token of the type is *indiscernible* from a constructed token. For example, for the finite type $\mathtt{2}$ the Indiscernibility Principle (corresponding to the above Uniqueness Principle) says

$$\prod_{\mathtt{x}:\mathtt{2}} \mathtt{Indis}_{\mathtt{2}}(\mathtt{x}, \alpha) + \mathtt{Indis}_{\mathtt{2}}(\mathtt{x}, \beta).$$

To prove this we use the induction principle for $\mathtt{2}$, which says that to derive a dependent function of type $\prod_{\mathtt{x}:\mathtt{2}} \mathtt{T}(\mathtt{x})$ (for any predicate $\mathtt{T} : \mathtt{2} \to \mathcal{U}$) it is sufficient to give its values at α and β. Since we trivially have tokens of $\mathtt{Indis}_{\mathtt{2}}(\alpha, \alpha)$ and $\mathtt{Indis}_{\mathtt{2}}(\beta, \beta)$ the Indiscernibility Principle for $\mathtt{2}$ follows immediately.

Thus, when given an arbitrary token $\mathtt{x} : \mathtt{2}$ to which $\mathtt{P}$ is to be applied, we may apply the dependent function witnessing the above Indiscernibility Principle to this token, thus yielding either a token of $\mathtt{Indis}_{\mathtt{2}}(\mathtt{x}, \alpha)$ or a token of $\mathtt{Indis}_{\mathtt{2}}(\mathtt{x}, \beta)$. From the resulting token we extract either a function of type $\mathtt{P}(\alpha) \to \mathtt{P}(\mathtt{x})$ or a function of type $\mathtt{P}(\beta) \to \mathtt{P}(\mathtt{x})$. Thus the Indiscernibility Principle ensures that function definitions in the above style are indeed exhaustive.[35]

In the next section we give two arguments for the claim that, regardless of the form or status of PII in HoTT, identity types must be taken as primitive components of the theory.

9. PROBLEMS WITH IDENTITY-AS-INDISCERNIBILITY

In this section we examine the three options outlined in Section 5 and find that none of them is successful. We thereby provide a counterargument to the idea that "identity types" in HoTT are really describing indiscernibility.

9.1. Recovering the Behaviour of `Id`

We begin with the second option from Section 5: taking the fragment of the language of HoTT without identity types (which in [Ladyman and Presnell, 2015] was called 'HoTT$^-$') and defining `Indis` in it instead of adding identity types to it. In order to make this system into a suitable substitute for HoTT one

[35] Of course, these derivations of Uniqueness Principles and Indiscernibility Principles are to some extent circular, since they use functions constructed according to the rules in order to justify the claim that these same rules are adequate (*i.e.*, exhaustive). It is better to understand the rules for function definition and the Uniqueness/Indiscernibility Principles not as two separate claims, the one justifying the other, but rather as two separate expressions of the grounding of identity and diversity facts in the theory. We discuss this further in Section 10.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

would have to show (at minimum) that `Indis` satisfied the defining properties of the identity type in HoTT.

First, we immediately recover (the analogue of) the token constructor `refl`, since indiscernibility is reflexive. Specifically, for any $\mathtt{a} : \mathtt{A}$ we have the function $\mathtt{refl}'_{\mathtt{a}} : \mathtt{Indis}_{\mathtt{A}}(\mathtt{a}, \mathtt{a})$ defined as

$$\mathtt{refl}'_{\mathtt{a}}(\mathtt{P}) :\equiv (1_{\mathtt{P}(\mathtt{a})}, 1_{\mathtt{P}(\mathtt{a})}),$$

i.e., the function that maps any predicate $\mathtt{P} : \mathtt{A} \to \mathcal{U}$ to the trivial demonstration that $\mathtt{P}(\mathtt{a}) \leftrightarrow \mathtt{P}(\mathtt{a})$.

The other part of the definition of `Id` that must be recovered is *path induction* (Section 2.3). For any type `A` and any token $\mathtt{a} : \mathtt{A}$, define

$$\mathtt{E}'(\mathtt{a}) :\equiv \sum_{\mathtt{x}:\mathtt{A}} \mathtt{Indis}_{\mathtt{A}}(\mathtt{a}, \mathtt{x})$$

whose tokens are dependent pairs $\langle \mathtt{b}, \mathtt{i} \rangle$, where $\mathtt{b} : \mathtt{A}$ and $\mathtt{i} : \mathtt{Indis}_{\mathtt{A}}(\mathtt{a}, \mathtt{b})$. Then the counterpart to (based) path induction says that for any predicate `K` that can be asserted of tokens of $\mathtt{E}'(\mathtt{a})$, if `K` holds of $\langle \mathtt{a}, \mathtt{refl}'_{\mathtt{a}} \rangle$ then it holds of *all* tokens $\langle \mathtt{b}, \mathtt{i} \rangle : \mathtt{E}'(\mathtt{a})$. More formally, it says that for any such predicate `K` there is a function

$$\mathtt{K}(\langle \mathtt{a}, \mathtt{refl}'_{\mathtt{a}} \rangle) \to \prod_{\langle \mathtt{b}, \mathtt{i} \rangle : \mathtt{E}'(\mathtt{a})} \mathtt{K}(\langle \mathtt{b}, \mathtt{i} \rangle).$$

In HoTT, path induction was incorporated as part of the definition of identity types, and so we were free to stipulate that it holds for identity; the remaining task was then to justify this stipulation on pre-mathematical grounds, which we did in [Ladyman and Presnell, 2015]. In contrast, in the present case the type `Indis` is already defined (as in Section 6.4) and so we are not free to stipulate additionally that it satisfies the above property. Rather, whether or not the above property holds of `Indis` is a matter to be proved as a theorem; as yet we have been unable to prove such a theorem or to find a counterexample.[36]

Alternatively, one might resort to imposing the above statement as a new axiom of the theory by assuming the existence of a token of the appropriate type (in the same way that Univalence is assumed as an axiom in HoTT). However, while it may be possible to give a pre-mathematical justification for such an axiom, along the lines of the arguments given in [Ladyman and Presnell, 2015], we consider this approach to be an inadequate substitute.

[36] Most obvious approaches to constructing counterexamples (*i.e.*, properties that hold of $1_{\mathtt{P}(\mathtt{a})} : \mathtt{P}(\mathtt{a}) \to \mathtt{P}(\mathtt{a})$ but not of $\mathtt{i}(\mathtt{P}) : \mathtt{P}(\mathtt{a}) \to \mathtt{P}(\mathtt{b})$) end up producing ways to discern `a` from `b`, in contradiction to the premise that `a` and `b` are indiscernible. However, because of the gap between `Indis` and `¬Dis` discussed in Section 6.5 it does not appear that this observation can be extended to give a proof of the required theorem.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

Versions of this problem affect all three options considered in Section 5. On each of those approaches we need to give a justification for the rules governing identity which should derive from the understanding of "identity" as indiscernibility.

9.2. Universes

While the problem of recovering path induction affects all three approaches, there is another substantial difference between identity and indiscernibility that obstructs their formal unification and therefore rules out options 2 and 3.

The definition of `Indis`,

$$\mathtt{Indis}_{\mathtt{A}}(\mathtt{a},\mathtt{b}) :\equiv \prod_{\mathtt{P}:\mathtt{A}\to\mathcal{U}} \mathtt{P}(\mathtt{a}) \leftrightarrow \mathtt{P}(\mathtt{b}),$$

involves a quantification over all predicates defined on type `A`, to express the idea that all properties definable on tokens of `A` are had in common by `a` and `b`. The type of the predicates that we quantify over is written as $\mathtt{A} \to \mathcal{U}$, where $\mathcal{U}$ is the *universe* — roughly, the type whose tokens are all the types under consideration.[37]

However, there is no single universe that contains all the types that can be defined within HoTT; it is a defining characteristic of universes that for any universe $\mathcal{U}$ we can find a "larger" universe $\mathcal{U}'$ distinct from $\mathcal{U}$, and in particular having $\mathcal{U} : \mathcal{U}'$. Thus we cannot pick in advance a single universe to be the domain of all our mathematical reasoning: at any point we may shift our attention from working in universe $\mathcal{U}$ to working in some larger universe $\mathcal{V}$.

We must therefore consider the possibility of different versions of `Indis` that quantify over predicates mapping into different universes. That is, rather than saying that two tokens are simply "indiscernible", we should say, for some universe $\mathcal{V}$, that they are "indiscernible relative to universe $\mathcal{V}$" (*i.e.*, they share all properties that map into $\mathcal{V}$), defined as:

$$\mathtt{Indis}^{\mathcal{V}}_{\mathtt{A}}(\mathtt{a},\mathtt{b}) :\equiv \prod_{\mathtt{P}:\mathtt{A}\to\mathcal{V}} \mathtt{P}(\mathtt{a}) \leftrightarrow \mathtt{P}(\mathtt{b}).$$

Since universes are cumulative, all properties definable in $\mathcal{U}$ are also present in any larger universe $\mathcal{V}$. Thus if `a` and `b` are discerned by $\mathtt{P} : \mathtt{A} \to \mathcal{U}$ then they remain discernible when we move to larger universe $\mathcal{V}$. However, it is also possible that new properties may become definable when we move from $\mathcal{U}$ to $\mathcal{V}$. We cannot rule out that two tokens that were indiscernible with respect to all properties definable in $\mathcal{U}$ may be discerned by a new property definable in $\mathcal{V}$. Thus it is not clear whether the fact that two tokens are indiscernible relative to $\mathcal{U}$ implies that they are also indiscernible relative to any larger universe $\mathcal{V}$.

Thus, whereas identity is a single relation that may hold between a pair of tokens, indiscernibility is a family of relations indexed to universes. This

[37] For a more detailed discussion of universes in HoTT, including an explanation of their most important properties, see [Ladyman and Presnell, forthcoming].

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

difference in character between identity and indiscernibility has important consequences that are explored in the next subsection.

9.2.1. Ramifications

Many important notions in HoTT are defined in terms of identity. For example, the definition of uniqueness is given in terms of identity: we say that some type $\mathtt{X}$ is uniquely inhabited by exhibiting a token $\mathtt{x} : \mathtt{X}$ and saying that all tokens of $\mathtt{X}$ are identical to $\mathtt{x}$:

$$\sum_{\mathtt{x}:\mathtt{X}} \prod_{\mathtt{y}:\mathtt{X}} \mathtt{Id}_{\mathtt{X}}(\mathtt{x}, \mathtt{y}).$$

Uniqueness is of course an important notion in mathematics, as when we say for example that a group element has a unique inverse.[38]

Another important example is the definition of a "mere proposition" as a type that has *at most one token*, which is expressed by saying that any tokens of that type are identical to one another. These types are central to the recovery in HoTT of traditional propositional logic [HoTT Book, Chapters 3.3 and 3.7]. Furthermore, types are classified into levels using these notions and others such as "set" and "groupoid" that similarly are defined in terms of identity.

Having identity as a primitive of the language of HoTT means that notions such as *uniqueness* and *mere proposition* can be given a singular definition that is not indexed to universes. Thus, if some type $\mathtt{X}$ is a mere proposition, or is uniquely inhabited, then this is a property of that type itself and holds of it regardless of which universe we work in.

If, on the other hand, we were to replace identity with a notion such as indiscernibility that is indexed to universes in the manner described in the previous subsection, then this relativisation would infect any notions defined in terms of it. For example, if we defined a notion of "uniqueness" in terms of indiscernibility then, as argued above, this would have to be "indiscernibility relative to some universe $\mathcal{V}$", and so the derived notion of uniqueness would likewise be "relative to universe $\mathcal{V}$"

$$\sum_{\mathtt{x}:\mathtt{X}} \prod_{\mathtt{y}:\mathtt{X}} \mathtt{Indis}^{\mathcal{V}}_{\mathtt{X}}(\mathtt{x}, \mathtt{y}).$$

Since we cannot guarantee that tokens indiscernible relative to $\mathcal{V}$ are also indiscernible relative to some larger $\mathcal{V}'$, consequently given a type that is uniquely inhabited relative to $\mathcal{V}$ it does not follow that it is uniquely inhabited relative to $\mathcal{V}'$. The same problem applies for all other notions defined in terms of identity.

This issue is particularly pressing in the case of *equivalence*, since the Univalence Axiom (Section 4.2) is defined in terms of equivalence. Recall

[38] In the HoTT Book this notion is called "contractability", which is a central notion in homotopy theory. Uniqueness is also needed for our justification of path induction via what we call the "Uniqueness Principle for identity types" [Ladyman and Presnell, 2015, Section 6.2].

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

that the definition of equivalence given in Section 4.1 involves identifications $(\mathtt{f} \circ \mathtt{h}_1)(\mathtt{b}) = \mathtt{b}$ and $(\mathtt{h}_2 \circ \mathtt{f})(\mathtt{a}) = \mathtt{a}$. Hence, if identity is indexed to universes then so is this definition of equivalence (and likewise isomorphism).

9.2.2. Typical Ambiguity

The problem identified above stems from our inability to express precisely the intended notion of indiscernibility. The idea we want to convey (especially if we are thinking of indiscernibility as a substitute for identity) is that tokens $\mathtt{a}$ and $\mathtt{b}$ of $\mathtt{A}$ are indiscernible by all properties whatsoever, *i.e.*, that $\mathtt{P}(\mathtt{a}) \leftrightarrow \mathtt{P}(\mathtt{b})$ holds for every property $\mathtt{P}$ on $\mathtt{A}$, without restriction. The nearest we can get to this is to quantify within a universe $\mathcal{U}$, and so over the type of predicates $\mathtt{A} \to \mathcal{U}$. However, while this captures all the predicates on $\mathtt{A}$ that are definable in the universe in which we are presently working, it fails to account for the fact that we may at some point switch to a larger universe $\mathcal{V}$, after which it may be possible to introduce some new predicate $\mathtt{Q}$ for which $\mathtt{Q}(\mathtt{a}) \leftrightarrow \mathtt{Q}(\mathtt{b})$ fails to hold.

To overcome this, we may wish to extend the statement of indiscernibility to say something corresponding to "$\mathtt{P}(\mathtt{a}) \leftrightarrow \mathtt{P}(\mathtt{b})$ for all properties $\mathtt{P} : \mathtt{A} \to \mathcal{V}$, for *any* universe $\mathcal{V}$". However, in HoTT there is no type or universe of all universes (indeed, positing such a universe leads to inconsistency [Coquand, 1992]) over which we could quantify to express the idea that something holds "for all universes".

One remaining hope is that we could express our intended meaning by means of *typical ambiguity*. This involves taking "$\mathcal{U}$" not to denote one particular universe, but rather as a "dummy variable" that can stand for any universe [Shulman, 2012; Ladyman and Presnell, forthcoming, Section 2]. For example, the formal statement of the Univalence Axiom is to be understood in a typically ambiguous manner (see footnote 19). Typical ambiguity may be thought of as a schema, like the induction schema in first-order logic. Thus, although we cannot quantify over universes within the language, and so we cannot express all instantiations of Univalence in a single statement, we have a "template" into which any universe may be inserted to give an expression of the Univalence of that universe. Correspondingly, then, we might hope to give the definition of indiscernibility a typically ambiguous reading: when we define

$$\mathtt{Indis}_{\mathtt{A}}(\mathtt{a}, \mathtt{b}) :\equiv \prod_{\mathtt{P}:\mathtt{A}\to\mathcal{U}} \mathtt{P}(\mathtt{a}) \leftrightarrow \mathtt{P}(\mathtt{b})$$

we intend "$\mathcal{U}$" to be read as a dummy variable, and so the intended reading is that *whatever* universe $\mathcal{U}$ is introduced, $\mathtt{P}(\mathtt{a}) \leftrightarrow \mathtt{P}(\mathtt{b})$ holds for all properties $\mathtt{P}$ definable in that universe.

Unfortunately, while it is legitimate to give such a typically ambiguous reading to the axioms that define the theory itself, it is not possible to do the same for definitions within the theory. Any type defined in the theory must live in some particular universe, but if we try to force a typically ambiguous reading

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

of "$\mathcal{U}$" in the above definition of $\mathtt{Indis}_{\mathtt{A}}(\mathtt{a},\mathtt{b})$ then there is no determinate universe in which this type lives.[39] Hence, typical ambiguity does not provide a way out of this problem and so indiscernibility remains indexed to universes, with all the unwanted consequences indicated in the previous subsection.

10. CONCLUSION

We have argued that identity types cannot be reconstrued as indiscernibility types for two reasons: first, because we cannot recover for indiscernibility the basic features of identity (Section 9.1), and second because any formal definition of indiscernibility must be indexed to universes whereas identity types are not (Section 9.2).

We have seen that even without identity types in the language we can recover the expected facts about discernibility and indiscernibility of tokens. In particular, for any uniqueness principle in HoTT we expect in general to have a corresponding indiscernibility principle in HoTT^{-}. The proofs of these principles and the construction of characteristic predicates (where these are possible) indicates that the facts about identity and diversity of tokens are *built into* recursion/induction rules. In many cases, the definition of the recursion/induction rule for a type encodes the intention that all tokens of the type are internally identical to the output of a constructor, and the uniqueness or indiscernibility principles then reflect this intention into the language.[40]

Type definitions in HoTT involve both constructors and recursion/induction rules. Since the latter control what can be said in the theory about the tokens of a type, they enable us to control what tokens can be said to exist in the type (up to identity or indiscernibility in the type). Very roughly, we may say that the constructors set a lower bound on what tokens exist, while the recursion/induction rules set an upper bound on what tokens exist.

In the light of the above, it is clear that identity and diversity facts in HoTT are not grounded in qualitative facts, since all such facts are expressed in the form of predicates and the rules governing the definition of predicates (and functions more generally) already encode the identity and diversity facts. In this sense, identity and diversity facts are primitive, although they are often encoded in the recursion/induction rules.

In Section 3.4 we raised the question of whether internal identity in HoTT is worthy of the name. Clearly, there is the distinct notion of external identity which lacks the novel features of identity types, but this is not in itself grounds to deny that `Id` may also deserve to be called identity — this would simply be to rule out intensional theories *a priori*. Furthermore, whereas types in the theory correspond to propositions, external identities do not. Thus identity types are the only way that propositions expressing identities can be represented

[39] We are grateful to an anonymous referee for clarifying this issue.

[40] However, these principles do not rule out "exotic models" in which (from an external point of view) a type may have additional elements that go beyond the intended content of the type [Bauer, 2013].

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

within the theory. Moreover, they are essential in HoTT for the formulation of Univalence.

Multiple and higher identities are novel features of identity in HoTT, but they are not in contradiction to standard features of identity. The most obvious reconstrual of `Id` is as indiscernibility, but we have seen that this does not work. Hence we conclude that, as is standardly supposed in work on HoTT, "identity" types really do express identity. So, unavoidably, this potential new foundation for mathematics has at its heart a radical new way of thinking about identity, which throws up new questions to be studied and, potentially, unexpected new applications to the many philosophical problems involving identity.

APPENDIX: THE FORMULATION OF PII IN HoTT

Recall from Section 6.6 Leibniz's formulation of PII: 'it is not true that two substances may be exactly alike and differ only numerically, solo numero'. Fixing some type $\mathtt{A}$, the most direct translation of this is:

$$\neg \sum_{\mathtt{x},\mathtt{y}:\mathtt{A}} \big(\mathtt{Alike}(\mathtt{x},\mathtt{y}) \times \neg \mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})\big),$$

where $\mathtt{Alike}(\mathtt{x},\mathtt{y})$ expresses that $\mathtt{x}$ and $\mathtt{y}$ have all properties in common. Below we consider different definitions of `Alike`. First we consider different forms of PII that arise however $\mathtt{Alike}(\mathtt{x},\mathtt{y})$ is defined.[41]

Since $\neg \sum$ and $\prod \neg$ are constructively equivalent, we can rearrange the above to:

$$\prod_{\mathtt{x},\mathtt{y}:\mathtt{A}} \neg\big(\mathtt{Alike}(\mathtt{x},\mathtt{y}) \times \neg \mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})\big).$$

Arranged this way, the body of the quantified proposition has the form of one of the three (classically equivalent) variants of material conditional:

$$\neg\mathtt{A} + \mathtt{B} \;\vdash\; \mathtt{A} \to \mathtt{B} \;\vdash\; \neg(\mathtt{A} \times \neg\mathtt{B}).$$

Constructively both entailments are strictly one-way. (Note that $\neg(\mathtt{A} \times \neg\mathtt{B})$ is materially equivalent to $\mathtt{A} \to \neg\neg\mathtt{B}$; so we shall use the latter variant.)

Variants of PII

Each variant of the material conditional gives a corresponding variant of PII:

$$PII^{++} :\equiv \prod_{\mathtt{x},\mathtt{y}:\mathtt{A}} \big(\neg\mathtt{Alike}(\mathtt{x},\mathtt{y}) + \mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})\big)$$

$$PII^{+} :\equiv \prod_{\mathtt{x},\mathtt{y}:\mathtt{A}} \big(\mathtt{Alike}(\mathtt{x},\mathtt{y}) \to \mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})\big)$$

[41] Analogously different formulations of PII arise in any constructive setting making the below taxonomy applicable more widely. Of course all these formulations of PII are trivially true unless the scope of quantification over predicates is restricted somehow as discussed in Section 6.1.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

$$PII^{-} :\equiv \prod_{\mathtt{x},\mathtt{y}:\mathtt{A}} \big(\mathtt{Alike}(\mathtt{x},\mathtt{y}) \to \neg\neg \mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})\big).$$

PII^{++} seems too strong to be reasonable: it says that for any two tokens of $\mathtt{A}$ we know either that they are identical or that they are not alike. But constructively we can be in the situation where there are two contrary possibilities but we know neither of them. So we would not expect PII^{++} to hold for most types, and we will not consider it further for now.

Variants of Alike

For any given $\mathtt{x},\mathtt{y} : \mathtt{A}$, the type $\mathtt{Alike}(\mathtt{x},\mathtt{y})$ says that $\mathtt{x}$ and $\mathtt{y}$ share all their properties. Thus it involves a quantification over predicates on $\mathtt{A}$. As with PII this has multiple formulations corresponding to the different variants of the material conditional.

$$\begin{aligned}
\mathtt{Alike}^{++}(\mathtt{x},\mathtt{y}) &:\equiv \prod_{\mathtt{P}:\mathtt{A}\to\mathcal{U}} \big(\neg\mathtt{P}(\mathtt{x}) + \mathtt{P}(\mathtt{y})\big) \times \big(\neg\mathtt{P}(\mathtt{y}) + \mathtt{P}(\mathtt{x})\big)\\
\mathtt{Alike}^{+}(\mathtt{x},\mathtt{y}) &:\equiv \prod_{\mathtt{P}:\mathtt{A}\to\mathcal{U}} \big(\mathtt{P}(\mathtt{x}) \to \mathtt{P}(\mathtt{y})\big) \times \big(\mathtt{P}(\mathtt{y}) \to \mathtt{P}(\mathtt{x})\big)\\
\mathtt{Alike}^{-}(\mathtt{x},\mathtt{y}) &:\equiv \prod_{\mathtt{P}:\mathtt{A}\to\mathcal{U}} \neg\big(\mathtt{P}(\mathtt{x}) \times \neg\mathtt{P}(\mathtt{y})\big) \times \neg\big(\mathtt{P}(\mathtt{y}) \times \neg\mathtt{P}(\mathtt{x})\big).
\end{aligned}$$

By distributivity, $\mathtt{Alike}^{++}(\mathtt{x},\mathtt{y})$ can be reformulated as

$$\prod_{\mathtt{P}:\mathtt{A}\to\mathcal{U}} \big(\mathtt{P}(\mathtt{x}) \times \mathtt{P}(\mathtt{y})\big) + \big(\neg\mathtt{P}(\mathtt{x}) \times \neg\mathtt{P}(\mathtt{y})\big).$$

However, to capture the intended meaning of "alike" we must have $\mathtt{Alike}(\mathtt{x},\mathtt{x})$ for any $\mathtt{x} : \mathtt{A}$. Then $\mathtt{Alike}^{++}(\mathtt{x},\mathtt{x}) \equiv \prod_{\mathtt{P}:\mathtt{A}\to\mathcal{U}} \mathtt{P}(\mathtt{x}) + \neg\mathtt{P}(\mathtt{x})$ asserts that every predicate on $\mathtt{A}$ is decidable, which is too strong. We therefore set aside $\mathtt{Alike}^{++}$ and do not consider it further.

$\mathtt{Alike}^{+}(\mathtt{x},\mathtt{y})$ is just $\mathtt{Indis}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$ as defined in Section 6.4.

By a constructive de Morgan law, $\mathtt{Alike}^{-}(\mathtt{x},\mathtt{y})$ is just the negation of $\mathtt{Dis}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$ as defined in Section 6.3.

Four Forms of PII

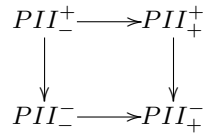
Putting these variants together, we therefore get four forms of PII:

$$\begin{aligned}
PII^{+}_{+} &:\equiv \mathtt{Alike}^{+}(\mathtt{x},\mathtt{y}) \to \mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y}) &&\equiv \mathtt{Indis}_{\mathtt{A}}(\mathtt{x},\mathtt{y}) \to \mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y}),\\
PII^{+}_{-} &:\equiv \mathtt{Alike}^{-}(\mathtt{x},\mathtt{y}) \to \mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y}) &&\equiv \neg\mathtt{Dis}_{\mathtt{A}}(\mathtt{x},\mathtt{y}) \to \mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y}),\\
PII^{-}_{+} &:\equiv \mathtt{Alike}^{+}(\mathtt{x},\mathtt{y}) \to \neg\neg\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y}) &&\equiv \mathtt{Indis}_{\mathtt{A}}(\mathtt{x},\mathtt{y}) \to \neg\neg\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y}),\\
PII^{-}_{-} &:\equiv \mathtt{Alike}^{-}(\mathtt{x},\mathtt{y}) \to \neg\neg\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y}) &&\equiv \neg\mathtt{Dis}_{\mathtt{A}}(\mathtt{x},\mathtt{y}) \to \neg\neg\mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y}).
\end{aligned}$$

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

10.1. Entailments

The following entailments follow immediately from the relative strengths of the variants of material conditional involved:



Since the conclusion of PII^-_+ and PII^-_- is the rather weak $\neg\neg \mathtt{Id}_{\mathtt{A}}(\mathtt{x},\mathtt{y})$, these ought to be acceptable even to the opponent of PII, but arguably only because they are too weak to be of interest. That leaves us with PII^+_+ and PII^+_-.

Finally, the principles that

(a) per the Curry-Howard correspondence, classical implication ($\Rightarrow$) corresponds to the existence of a function ($\rightarrow$); and
(b) in translating ideas from classical to constructive logic we should seek to avoid negation where possible

lead us to take PII^+_+ to be the right version of PII in a constructive setting. This corresponds to the natural reading of the Principle of Identity of Indiscernibles, namely that "any things that are indiscernible are identical".

REFERENCES

Awodey, Steve [2014]: 'Structuralism, invariance, and univalence', *Philosophia Mathematica* (3) **22**, 1–11.

Baez, John C., and James Dolan [1998]: 'Categorification', in Ezra Getzler and Mikhail Kapranov, eds, *Higher Category Theory*, pp. 1–36. Contemp. Math.; 230. Providence, Rhode Island: American Mathematical Society.

Bauer, Andrej [2013]: 'The elements of an inductive type', http://math.andrej.com/2013/08/28/the-elements-of-an-inductive-type Accessed August, 2016.

Caulton, Adam, and Jeremy Butterfield [2012] : 'On kinds of indiscernibility in logic and metaphysics', *The British Journal for the Philosophy of Science* **63**, 27–84.

Coquand, Thierry [1992]: 'The paradox of trees in type theory', *BIT Numerical Mathematics* **32**, 10–14.

Gambino, Nicola, Chris Kapulkin, and Peter LeFanu Lumsdaine [2011] : '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. http://www-home.math.uwo.ca/ kkapulki/notes/UA_implies_FE.pdf. Accessed August, 2016.

Geach, P.T. [1967]: 'Identity', *Review of Metaphysics* **21**, 3–12.

Hofmann, Martin, and Thomas Streicher [1998]: 'The groupoid interpretation of type theory', in *Twenty-five Years of Constructive Type Theory (Venice, 1995)*, pp. 83–111. Oxford Logic Guides; 36. Oxford University Press.

Ladyman, James, and Stuart Presnell [2014]: *A Primer on Homotopy Type Theory*. http://philsci-archive.pitt.edu/11157/. Accessed October, 2016.

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016

_____ [2015]: 'Identity in Homotopy Type Theory, Part I: The justification of path induction', *Philosophia Mathematica* (3) **23**, 386–406.

_____ [2016]: 'Does Homotopy Type Theory provide a foundation for mathematics?', *The British Journal for the Philosophy of Science*. DOI: 10.1093/bjps/axw006.

_____ [forthcoming]: 'Universes and univalence in homotopy type theory', *The Review of Symbolic Logic*.

Ladyman, James, Øystein Linnebo, and Richard Pettigrew [2012]: 'Identity and discernibility in philosophy and logic', *The Review of Symbolic Logic* **53**, 162–186.

Leibniz, Gottfried Wilhelm [1686/1962]: *Discourse on Metaphysics*. La Salle, Illinois: Open Court.

Leitgeb, Hannes, and James Ladyman [2008]: 'Criteria of identity and structuralist ontology', *Philosophia Mathematica* (3) **16**, 388–396.

Licata, D. R., and G. Brunerie [2015]: 'A cubical approach to synthetic homotopy theory', in *Logic in Computer Science (LICS), 2015 30th Annual ACM/IEEE Symposium*, pp. 92–103. Washington, D.C.: IEEE Computer Society. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7174873 Accessed August, 2016.

Martin-Löf, Per [1974]: 'An intuitionistic theory of types: Predicative part', in H.E. Rose and J.C. Shepherdson, eds, *Logic Colloquium '73. Proceedings of the Logic Colloquium, Bristol, July 1973*. pp. 73–118. Studies in Logic and the Foundations of Mathematics; 80. Amsterdam: North-Holland.

McTaggart, J.M.E. [1921]: *The Nature of Existence*. Cambridge Paperback Library. Cambridge University Press.

nLab [2014]: 'Homotopy type system'. http://ncatlab.org/homotopytypetheory/show/Homotopy+Type+System. Accessed August, 2016.

Quine, W.V. [1976]: 'Grades of discriminability', *Journal of Philosophy* **73**, 113–116.

Resnik, Michael D. [1997]: *Mathematics as a Science of Patterns*. Oxford: Clarendon Press.

Shapiro, Stewart [2008]: 'Identity, indiscernibility, and *ante rem* structuralism: The tale of i and $-i$', *Philosophia Mathematica* (3) **16**, 285–309.

Shulman, Michael [2012]: 'Universe polymorphism and typical ambiguity', Article at The n-Category Café, December 2012. https://golem.ph.utexas.edu/category/2012/12/universe_polymorphism_and_typi.html. Accessed August, 2016.

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

Voevodsky, Vladimir [2013]: 'A type system with two kinds of identity types', https://uf-ias-2012.wikispaces.com/file/view/HTS_slides.pdf. Accessed August, 2016.

doi: 10.1093/philmat/nkw023

Downloaded from http://philmat.oxfordjournals.org/ at University of Newcastle on December 21, 2016