Justifying Path Induction

An Inferentialist Analysis of Identity Elimination in Homotopy Type Theory

Patrick Walsh

A thesis presented for the degree of
Master of Science in Logic, Computation, and Methodology

Department of Philosophy
Carnegie Mellon University
Pittsburgh, Pennsylvania
United States
July 30, 2015

Justifying Path Induction

An Inferentialist Analysis of Identity Elimination in Homotopy Type Theory

Patrick Walsh

Abstract

Homotopy Type Theory (HoTT) has recently been challenged as to its foundational status. James Ladyman and Stuart Presnell have contested that HoTT's presentation is not sufficiently 'pre-mathematical'. I argue that indeed the presentation is not pre-mathematical and that's a good thing. I give a *mathematical* justification for the rule of inference in HoTT they find particularly problematic: Path Induction. I provide a categorical justification of both the internal structure and external representation of Path Induction. This justification, although mathematical, has its philosophic roots in philosophy of logic and language. Particularly, I draw from Prawitz style *inferentialism* and the notion of *Harmony* championed for example by Michael Dummett. This constellation of ideas is unified through categorical logic. In the end, I hope to have justified Path Induction in response to valid philosophical concerns.

Acknowledgements

I want to thank my adviser Steve Awodey for his direction and especially his help on the final proof of this paper. I am also deeply grateful for Wilfried Sieg's comments that often had me thinking hard about what I meant and what I wanted to say. To all my lovely and supportive friends in and out of the department, thank you so much. I couldn't have done it without lots of help and kind words.

Contents

1 Introduction **4**

2 Preliminaries **10**
2.1 Intuitionism and Type Theory 10
2.1.1 Proof Relevance 10
2.1.2 Type theory and Propositions 12
2.2 Martin-Löf Type Theory 13
2.2.1 Propositional and Definitional Equality 16
2.2.2 Identity Types 17

3 Ladyman, Presnell, and Foundations **23**
3.1 The Argument 24
3.2 The Criterion for Justification 27

4 Justifying Rules of Inference **33**
4.1 History 33
4.2 Dummett 38
4.3 An Inferentialist Treatment 40
4.4 Brandom's Inferentialism 43
4.5 Meaning and Truth 47

5 Harmony **50**
5.1 Conceptions of Harmony 51
5.2 Adjoints 57

6 Lawvere's Law **59**
6.1 Hyperdoctrines 61
6.2 Equality in a hyperdoctrine 69
6.3 Equivalent Identities 74

7 Conclusion **78**

1 Introduction

Justifying our rules of inference, in almost any context, is a difficult matter. The justification is necessarily more abstract than the language in which we use the rules of inference. It seems, then, that the justification for our rules of inference requires rules of inference in the meta-language governing this justification. This can continue since we need to justify the means by which the rules of inference are justified. But it is precisely this task of justifying a rule of inference that I take up in this paper. I hope to give a motivated and rigourous justification for a particularly subtle rule of inference: *Path Induction*. We might view my project as part of the new research program of Homotopy Type Theory (HoTT), but it will also engage with the philosophy of logic and mathematics. I would say that my work is *about* the HoTT project, instead of *in* it. The distinction is not all that important but I do want to indicate that readers are only expected to have an informal understanding of type theory.

Homotopy Type Theory is a new branch of mathematics that is known for combining apparently disparate fields. The homotopic aspect comes out in interpreting types of the type theory as spaces and terms of those types as points of the space. Type theory does not usually include such mathematical interpretations, as it is traditionally seen as a computational formalism. However, intensional type theory, which I will explain later, was known to exhibit a groupoid structure when included with dependent identity types [58]. Later, the identifications of terms were interpreted as *paths*, higher-dimensional objects in the space, which was first shown to be a sound interpretation in [8]. With such important proof assistants like `Coq` and `Agda` being based on this intensional type theory, this work started to influence how type theory was understood even among computer scientists.

An important and famous addition to this growing connection was Voevodsky's *Univalence Axiom*. Perhaps the most discussed consequence of this axiom is that isomorphic structures are identified. The strong structuralist flavor of this axiom, among other features, has given *Univalent Foundations* its status as a putative foundation. The combination of HoTT and univalence has thus led to a lively and quickly growing research program. Though this feature is the most discussed, the axiom redefines the relationship between equivalence and identity. Simply from

this very brief and incomplete history of the HoTT program we can already see the importance of the mathematical notion of *identity*. Indeed, considering the identity types as *path objects*, a homotopic notion, gives HoTT part of its rich structure. In this paper, I consider the more philosophical side of identity in Homotopy Type Theory.

This project is motivated by a challenge forwarded by James Ladyman and Stuart Presnell in [70]. They claim that the justification for an identity rule in HoTT is lacking a necessary justificatory feature. In contrast to the explicitly homotopical presentation given in the HoTT Book [155], Ladyman and Presnell require that the justification be 'pre-mathematical'; this is the feature they argue is lacking in the identity rules. They supply their own justification for the identity rules — where the contentious part is called *Path Induction* — that supposedly does not rest on higher mathematical notions like homotopy. Although they are motivated by the mathematical *presentation* of the identity rules in the HoTT Book, they take it as a sign that the *justification* of the identity rules is also lacking a non-mathematical character.

I think this challenge is appropriate. Path Induction is surely a subtle and powerful rule in HoTT and its nature might need both explanation and justification. But I do not think that the justification given in Ladyman and Presnell's [70] is sufficient or well motivated. However, my project is focused on creating an alternative, not on criticizing the program advanced by Ladyman and Presnell. I do forward some objections to the Ladyman and Presnell account in Chapter 3, but this should only serve to motivate why I think the project is still to be completed. The main argument I make is against their requirement that the rules of inference, like Path Induction, be pre-mathematically justified. My account will certainly not be pre-mathematical and I would even argue that neither is their account.

To create my positive program for justifying the rules of HoTT, specifically Path Induction, I need a framework in which to work. Part of the program of justifying rules of inference requires that we know exactly what it is we are justifying. Some account of meaning and content will be needed to fix this concept of identity. Accordingly, I will be using a brand of *inferentialism* as my framework at the philosophical level. Inferentialism, broadly, is the thesis that for any concept, the content of that concept is determined by the inference moves to and from that concept.[1] Inferentialism is a wide-ranging program that can cover all of natural language; I do not take this position here. Instead, I use inferentialism in a narrow way to express a certain pragmatism in the formal languages of logic and

[1] To be careful here, I would point out that inference rules cannot be to and from *concepts* since concepts are not propositions. So we consider the inference rules to and from a concept's *ascription* to determine the content of that concept.

mathematics. That is, inferentialism, as I'll use it, only says that mathematical and logical concepts are determined by their explicit inferential relations. The notion of pragmatism, then, means that it is the inferential *use* of the concept that determines what that concept is.

It is not obvious that this change of context from the natural to the mathematical will help in getting inferentialism off the ground until we have a way to mathematically talk about this conceptual content. How do we characterize the "inferential relations" of a logical or mathematical concept in a sufficiently general way? The answer is in the framework of category theory, or categorical logic. Category theory, I'd argue, is fundamentally an inferentialist language. Indeed, a category theorist's slogan "It's the arrows that really matter" [2, p. 8] can be interpreted as saying precisely that the arrows are sufficient to determine the 'nature' of the objects. The arrows of category theory, in our exposition, will be interpreted as deductions, or inferences, giving us a strong inferentialist reading of the slogan. Crucial features like the Yoneda Lemma and the universal characterizations of objects enforce my contention that category theory is a framework in which to make inferentialism precise, at least for our purposes. It will also be important that categorical logic, through the *hyperdoctrines* that we will use, covers several different logical frameworks, including the classical one. This generality will give credibility to one of the major purposes of the paper: showing that the identity rules in HoTT exhibit the 'universal character' of identity. Here, I mean to use the word "universal" to indicate both an independence from questions of which logic is ideal but also to reinforce the tie to the 'universal characterizations' in category theory.

Thus by using an inferentialist thesis we can characterize the concepts of HoTT by considering the inferential rules that govern them. Further, we consider these inference rules using categorical logic. And so, using our philosophical program of inferentialism and the mathematical language of category theory we can determine when a set of inference rules appropriately fixes the meaning of a concept. With this, we arrive at a categorical notion of both identity and of what constitutes a 'good' rule of inference. I claim that these together will justify Path Induction, as a rule of inference governing identity.

Per Martin-Löf, whose eponymous type theory HoTT is based on,[2] often gave what are referred to as 'meaning explanations' for most of the types he used as well as the notions of judgment, and context [91]. The identity types that we present in this paper have not received such a meaning explanation. Usually these

[2]It is interesting that in the progress of HoTT, researchers refer to the many interpretations or perspectives of HoTT. We might, for example, discuss the 'Book HoTT' or the recent work in semantics or 'cubical type theory'. This isn't so much a divide in the research community as a pluralism with respect to perspective and HoTT's ability to bear such pluralism. However, some have desired to jettison the idea that HoTT is based on Martin-Löf Type Theory.

explanations take the form of appeals to intuition and characterizations in terms of intended interpretations. My meaning explanation of the identity type will not be Martin-Löf's, but it is intended to be of a piece, although significantly elaborated, with his meaning explanations. The (mathematical) intuition and motivation behind the formulation of the identity rules in Martin-Löf Type Theory will be given in Section 2.2.2.

The difference between explanation and justification needs to be looked at closely if we are to give such a meaning explanation. When Martin-Löf gives meaning explanations of the types, this is something like a 'reality check' where we confirm that the rules of that type do what we want them to. But how is this different from a *justification* of the rule of inference? I believe that justifications are partly explanations but they also include a pragmatic matching of ends and means. Of course we might not want to generalize this to every sort of explanation, but I view the meaning explanations given by Martin-Löf as partial justifications. By this I mean that there is a dimension of the explanation that speaks to the legitimacy of the rule of inference. We might view such an explanation as a necessary condition for the justification of the rule, but surely, it might be supposed, being able to give an interpretation of the rule is not sufficient to justify it. I believe that the purpose of an explanation is to match expectations and desires with the actual logic or type theory we have created. Justifications, on the other hand, must also show why we should want these features at all, requiring a slightly elevated analysis of the position the logical system occupies in our conceptual framework. The way that I approach *justifying* Path Induction is by explaining how it operates and then showing why we want it to operate the way it does. I believe this pragmatic approach to justifying is an eminently reasonable one and throughout this paper I will often oppose foundationalist requirements for justification, like that of Ladyman and Presnell. Where they require the concept of identity to be non-mathematical, I do not. To put it another way, the pragmatic matching of norms to actualities need not be constrained to norms of intuition or the pre-mathematical; they may also be properly mathematical.

I now outline the structure of the paper. The next chapter, Chapter 2, will go over some of the technical concepts and techniques that we need. The first section is useful for those who are not familiar with 'propositions as types' in type theory or interpretations of intuitionistic logic. Section 2.2 goes into more technical detail about the type theory that HoTT is based on. Most importantly, the way that Path Induction and the identity types are presented in Martin-Löf type theory is explained in this chapter. Getting a feel for the rule and the way it works is crucial for our goal of justifying the rule. We'll also give several perspectives on the central rule of this paper: Identity Elimination (Path Induction). This chapter should equip the reader with the tools necessary to consider whether the

justification of Path Induction given later is sufficient.

I then summarize and provide a rebuttal of the argument made by Ladyman and Presnell. This will occupy us for Chapter 3. Again, I want to emphasize that my project is mainly positive, not negative. So, although I take the account given in [70] to be incorrect, I focus more on how to achieve our common goal through different means. The motivation of the project is due to the work by Ladyman and Presnell and I hope to recognize this fact by taking their concerns seriously. This also helps in situating my project with respect to theirs.

Chapter 4 begins the positive account for justifying rules of inference. Here, I outline the history of justifying rules of inference and how my project fits in with this interesting historical trajectory. I will ultimately land on an account similar to those of both Prawitz and Dummett. This account will allow the separation of the justification of a rule of inference into two parts: internal and external. The internal question is that of *Harmony* which says that the introduction and elimination rules are appropriately balanced with each other. In other words, regardless of whether the rule represents the correct notion or whether we want to allow it into our foundation, the rule must be internally coherent in a relevant way. This is what we call Harmony. The external question, on the other hand, asks whether the rule 'captures' the concept it purports to. This will require an independent characterization of the concept and some way to compare it with the rule. But the independent characterization of a mathematical concept need not be non-mathematical. The independence simply comes by considering a different corner of mathematics. When we justify the identity rules of HoTT, we will not be comparing it to the conception of identity in HoTT, of course. Instead, we will use Lawvere's notion of *hyperdoctrines* [77] to define mathematical identity in a general context, including proof-relevant and classical logics. Then we give the central consequences of this definition and show that HoTT's identity type has the same characteristics.

I then explain the inferentialist framework behind the program to come; in Section 4.3, I describe what an inferentialist framework requires and implies. I restrict my discussion there to inferentialism in mathematics. This could be markedly different from an inferentialist account of natural language. For instance, I will not need language entrance and exit moves in my inferentialism[3]— although I will comment on plausible formal language entrance and exit moves. In essence, this chapter attempts to give philosophical significance to the more mathematical chapters to follow. It will also, I think, frame the discussion on Harmony appropriately.

We are now in a position to consider each facet of a justification individually in the inferential framework of category theory. Chapter 5 covers the internal question

[3]These are the sorts of terms that Sellars uses to describe a certain kind of inference rule that we wouldn't represent with logical inference rules. More will be said of this in 4.3.

of justification. Using a categorical characterization of the relevant concepts as adjoints, I explain what Harmony is, this balance between the introduction and elimination rules of a concept. This conception will unify the many conceptions in the literature of both *harmony* and the dual notion of *stability*. The main thrust will be to argue that concepts determined by adjoints are harmonious, that is, appropriately balanced. I then work through an example of showing a concept's rules to be harmonious: the case of conjunction. This will lead us to showing that the identity rules of Martin-Löf type theory — and thus HoTT — are indeed harmonious. This will finish the internal justification of Path Induction and the other identity rules.

I then move to the external question of showing that Path Induction and the other identity rules capture the inferentially determined concept of identity in Chapter 6. In a sense, this is the climax of the paper, giving justification for Path Induction. I replace the traditional characterization of identity due to Leibniz with a categorical characterization of identity. Instead of Leibniz's laws of identity that focus on properties, I offer what I call "Lawvere's Law" which essentially states that identity is the initial reflexive relation. This shift from properties to relations will provide us with the leverage to show that identity is uniquely captured in Lawvere's Law. Given this categorical description of identity, I show that it represents the identity rules of Martin-Löf type theory and vice versa. And so Path Induction will correctly capture the notion of identity. This will constitute the external justification of Path Induction, concluding the paper.

2 Preliminaries

2.1 Intuitionism and Type Theory

In this section, I'll cover some important aspects of HoTT that will inform the way we approach answering the main questions of this paper. Two related notions that will underly everything in the type theory used are those of *proof relevance* and *propositions as types*. These together allow us to interpret intuitionistic logic in the type theory. First we'll discuss the general features of proof relevance and some context in which we find the concept.

2.1.1 Proof Relevance

Proof relevance is an important notion in what is normally called the BHK-interpretation of intuitionistic logic.[1] There are various ways of expressing this interpretation. A simple one for a proposition P is the following:

P is true if and only if it has a proof.

This identification between truth and provability is at the heart of intuitionism and constructivism as well as many other systems.[2] Of course, in light of Gödel's incompleteness theorem, we do not say provability is restricted to any particular formal system [91, p. 11].

This notion of proof relevance needs a little further explication. The "relevance" indicates that the proof is relevant to reasoning about the proposition it proves. A sort of inseparability between proof and proposition is required in proof relevant logics like ours. Expanding on this, I will give the usual presentation of the logical constants under the BHK-interpretation. This way, we will become acquainted

[1] Which can refer to the original Brouwer-Heyting-Kreisel in [152, p. 977] or the newer and more common meaning as the Brouwer-Heyting-Kolmogorov interpretation of intuitionism from [153, p. 10]. We will be using this later, more common, meaning.

[2] I argue later that closing the distance between proof and truth is related to the inferentialist program of closing the distance between inference and meaning. This is due to the proof/inference and truth/meaning connections that will explored in Chapter 4.

with the pattern of how we make concepts proof relevant. This is extended to all propositions in HoTT.

Table 1: The BHK-interpretation of propositions

A proof of	is
$\bot$	nonexistent.
$A \,\&\, B$	a pair (a, b) where a is a proof of A and b is a proof of B.
$A \vee B$	$i(a)$ where a is a proof of A **or** $j(b)$ where b is a proof of B.
$A \to B$	a method b that takes any proof a of A to a proof $b(a)$ of B.
$\forall x\, B(x)$	a method b that takes any individual c to a proof $b(c)$ of B.
$\exists x\, B(x)$	a pair (c, p) where a is an individual and p is a proof of $B(c)$.

The terms $i(a)$ and $j(b)$ indicate which disjunct is true, giving us the intuitionistic disjunction. Allowing ambiguity here about what sort of proofs these lowercase letters are, we notice that the proofs themselves are objects. This is intentionally reminiscent of proof-theoretic methodology. In our type theory, we maintain this ambiguity of which logic we're using and express that a is a proof of a proposition A by saying that a is a term of type A.

As a helpful contrast to MLTT and HoTT, classical mathematics is mostly proof *irrelevant*. A proof is usually a mechanism for establishing the binary truth value of a statement. For the BHK-interpretation of intuitionistic logic, proofs are mathematical objects in their own right. This allows transformations and something we might call 'proof tracking' rather than 'truth tracking' — if truth here is conceived of classically.[3] Proof theory straddles the difference between proof-relevant and proof-irrelevant logics. For example, in many sequent styles of natural deduction we can say that the sequent $\Gamma_1 \vdash \Delta_1$ is derivable given that $\Gamma_2 \vdash \Delta_2$ is derivable. But we can also say that given a derivation $\mathcal{D}$ of $\Gamma_1 \vdash \Delta_1$, by operating on it in such and such a way, we can arrive at a derivation of $\Gamma_2 \vdash \Delta_2$, where "such and such a way" is replaced by proof theoretic manipulations.

The most important notion in proof relevance is that to assert a mathematical proposition is to provide a proof of it. This is in contrast to classical mathematics which only presents the proposition, not necessarily with the proof in tow. We see this mostly with well-established theorems. We don't prove Euclidean facts by appeal to the Pythagorean Theorem *and* its proof; we usually just appeal to the theorem itself. Proof relevance doesn't allow the separation of assertion from proof.

[3] There is a tradition of taking truth to be defined in terms of proof. Philosophically aligned with the BHK-interpretation, but not restricted to intuitionism per se is that of *proof-theoretic semantics*. See for instance [113], [133], and [156].

2.1.2 Type theory and Propositions

Another important notion in HoTT is that of 'propositions as types'. We will say that the notation $a : A$ expresses the fact that a is a term of type A. When we view the type A as a proposition, this will be read as "a is a certificate of the truth of A". We don't always use the word "proof" to describe this relationship because terms need not have apparent structure, as proofs often are required to have. That is, these certificates are not sequences of formulas, nor are they trees; they're simply terms of a type. This means that propositions are types.[4] *A proposition is therefore the type of its proofs.* So when we want to prove $A \,\&\, B$, we need to give a pair (a, b) such that $a : A$ and $b : B$. This would amount to a proof of A and a proof of B, loosely speaking.

This is of a piece with the Curry-Howard correspondence [141], sometimes also called an 'isomorphism'. This correspondence is how we relate logical notions like quantifiers and connectives to computational notions like dependent Π types and function types. The relation between programs and proofs is exemplified in a type theoretic construction that includes logic, as HoTT does. This creates a close constellation of concepts with the BHK-interpretation and the Curry-Howard correspondence as well as the closely related concepts of realizability and what has been called the Curry-Howard-Lambek correspondence.[5] We might view HoTT as a generalization and formal framework where all these interpretations come together — the figure that gives the constellation unity.

As for the methods mentioned in the BHK-interpretation, these are lambda terms that express functions. So a method that takes a proof of A and returns a proof $b(a)$ of B will be verbosely expressed by lambda terms

$$\lambda(a : A).(b(a) : B)$$

The left side of the period expresses the input type and the right side expresses the resulting type. In other words, this lambda expression gives us an unnamed function $a \mapsto b(a)$ with appropriate types. For brevity, we will also represent this by $\lambda a.b(a)$, or even $\lambda a.b$, when the types are clear.

As alluded to above, a tight analogy can be drawn between natural deduction proofs and propositions as types. For instance, in proof theory, we can reason as follows: If we have a proof $\mathcal{D}$ of A, we can systematically create a proof of $A \vee B$ with some proposition B:

[4] We can think of propositions and types as being the same thing [91] or we might only require that some types are propositions [5]. We need not go into the homotopy levels of types or make a choice as to whether propositions and types are in general the same. But I will restrict my attention to those types that are propositions.

[5] This extension can be seen in Lambek's work in Cartesian closed categories and other categorical interpretations in [72].

$$\frac{\begin{array}{c}\mathcal{D}\\A\end{array}}{A \vee B}\ \vee\text{I}$$

This is in a sense an *operation* on the original $\mathcal{D}$ to a proof of $A \vee B$. Propositions as types makes this even more clear where the "disjunction introduction"[6] is expressed as

$$\frac{\mathcal{D} : A}{\texttt{inl}(\mathcal{D}) : A \vee B}$$

which might reasonably be read as "if we have a proof (term) $\mathcal{D}$ of A, then we can create a proof (term) of $A \vee B$ that we call $\texttt{inl}(\mathcal{D})$". We call it "`inl`" to indicate that the left disjunct has a proof which is turned into a proof of the disjunction. If the right disjunction is provable, we likewise would use $\texttt{inr}(\mathcal{D})$ for our proof of the disjunction. So `inl` takes proofs of A and gives a proof of $A \vee X$ for any X. This is precisely how we reasoned in the natural deduction case above. Operating on proofs as mathematical objects in their own right is common to proof theory and in proof relevant type theory.

And here we can neatly see the relation between the propositions as types and proof relevance. Proof relevance allows us to attach proofs to every proposition and 'propositions as types' allows us to represent this attachment as terms of types. Since there can be more than one proof of a proposition, we represent them as different terms of that type. In other words, proof relevance requires we keep track of proofs for every proposition. And by considering propositions as types and proofs as terms of those types, we have a formalism to accommodate the proof relevance.

This concludes the discussion of these two critical features of MLTT and HoTT. Later, when we consider the categorical models of these sorts of logics, we'll pay due attention to making sure the logic is proof-relevant. That is, the derivability relationship $\vdash$ that is modeled in category theory must admit many proofs, allowing for proof relevance.

2.2 Martin-Löf Type Theory

As briefly mentioned in the introduction, this paper is not properly within the technical framework of Homotopy Type Theory. In order to say anything substantive about HoTT, however, we have to discuss several technical parts of the framework. We will mostly restrict ourselves to the dependent type theory that

[6] This example isn't strictly correct as the disjunction should be a coproduct, but for illustration, I've kept it as logical disjunction. This is warranted by the correspondence between products and conjunctions, coproducts and disjunctions, functions and conditionals, etc.

underlies HoTT: that is, Martin-Löf Type Theory. The question of the foundational status of HoTT does require that we analyze the foundational status of Martin-Löf Type Theory (MLTT) and to be more precise, this is the true context for the paper. For this purpose, it is helpful to go over key aspects of MLTT that need analysis and specifically the components that Ladyman and Presnell find *prima facie* objectionable. Thus in this chapter I will describe Martin-Löf's type theory as laid out in [102] and some related materials.

First off, Martin-Löf type theory is a sort of *intensional* type theory. This is in contrast to *extensional* type theories. The word "intensional" refers to the nature of *types*. Whereas in set theory, an extensional framework, sets are completely determined by their elements, in intensional type theory, types are determined by how their objects are created — or introduced. This genetic way of distinguishing types has the consequence that each 'object' in this type theory can only be a member of one type.[7]

An example of the difference between extensional and intensional reasoning is the following: The set of human beings and the set of naturally featherless bipeds are extensionally equal. Every human being is a naturally featherless biped and vice versa. However, the meaning of 'human' and 'naturally featherless biped' are not the same. We do not treat these terms as synonymous. If a (currently) non-human mammal started exclusively locomoting bipedally, we'd feel no temptation to say that it's human.[8] This is an old point on the difference between sense and reference [47]. The types of MLTT are determined by their intension, accordingly.

Those accustomed to a set-theoretic framework like Zermelo-Fraenkel (ZF) set theory will notice the stark contrast intensional type theory presents. In ZF, the set membership relation "$\in$" is extensional. To avoid confusion, I will follow the authors of the HoTT Book [155] and not Martin-Löf's in [91] with the parallel notation for type-membership:

$$a : A$$

will denote that the *term* a is of *type* A. We say that a is a *canonical* term of type A if it is a direct consequence of one of A's *introduction rules*. This is what we call the rule for creating objects of type A. More will be said about this later.

Now we can see the crucial difference that Martin-Löf points to: A set — and other things he calls "categories" — is determined by what it takes to be a member. That is, what condition must an 'arbitrary element' satisfy in order to

[7] For an interesting parallel between this fact and Robert Brandom's theory, see Section 4.4.

[8] The difference here is relying somewhat on the intuition of the difference between a contingent and necessary truth where synonymy must be necessary. The meaningfulness of the distinction between intensional and extensional is not lost if we abandon this conception of synonymy. No more is needed for this intuition than that definitions are somewhat stable with new objects coming into our knowledge of the world.

be a member. This is usually the separation formula $\phi(x)$ in set builder notation like $\{x \in A \mid \phi(x)\}$. A type, however, is determined by two things:

1. Prescribing how canonical elements are *formed*;
2. Giving a criterion of (definitional) identity between canonical elements [91, p. 8]

The first of these is the introduction rule for a type. It allows us to form canonical terms of types. But in general, we don't immediately know whether all the terms of a type are canonical — since we are in an intuitionistic interpretation, we read "know" as "have a proof of". And secondly we impose an equivalence relation on these canonical terms so that we get definitional equality. One seemingly trivial thing this helps with is the identification of instances of the names of types below and above the line in inference rules and similar sorts of symbol tracking. After all, with a goal of mechanizability, we need a way to make sure proof assistants recognize that all instances of the type A in the formula

$$a : A \to \texttt{inl}(a) : A \vee B$$

are the same. This is of course a trivial matter for humans. We in fact find it hard *not* to see all instances of the same letter as representing the same thing. Clearly, if we cannot make this explicit, then we are without a hope of practical computability.

The proof relevance and propositions as types principles of this type theory provide the way we distinguish *how* we got a term $a : A$ from merely *that* we have a term of that type. Namely, continuing the example of disjunction introduction, we don't simply say $\dfrac{a : A}{b : A \vee B}$. This would not explicitly indicate that our proof of $A \vee B$ was derived from a proof of A. Indeed, intuitionistic disjunction requires knowing precisely which disjunct is true, hence the denial of a general law of excluded middle.[9] For this reason, we have canonical terms `inl` and `inr` that tell us which disjunct has a proof. This is what is meant by a 'genetic' way of distinguishing types.

Part of the interest of HoTT comes from the fact that proof relevance allows multiple proofs of a proposition. A canonical proof of a disjunction (a term of the form $\texttt{inl}(a)$ or $\texttt{inr}(b)$) would be proving a disjunct. But there may be other proofs of such a disjunction. As previously indicated, we can't immediately prove (finding a term of a specific type) that there are no other proofs of a disjunction.

[9] We can either deny the law of excluded middle by affirming its negation or simply deny the general validity of the law, allowing that it may be true in some cases. I will take the latter approach: we can introduce it when we want, but it is not part of the basic theory.

Intuitively, if $a:A$ and $b:B$ are distinct terms, then $\mathtt{inl}(a) : A \vee B$ and $\mathtt{inr}(b):A \vee B$ are as well. Clearly these introduction rules only give us sufficient conditions for defining terms of a type, but not the necessary ones. We haven't said anything about the necessity of these constructions of the terms in $A \vee B$.

2.2.1 Propositional and Definitional Equality

We now need to make precise what we mean when we specify *definitional equality* $\equiv$ required in (2) above and something we'll call *propositional equality* $=$. This separation of two sorts of equality has been the source of interesting questions in HoTT.

Definitional equality is an intensional equality. We would say that $s(s(0))$ has the *same meaning* as 2 in the natural numbers; indeed 2 is just an abbreviation of $s(s(0))$. Similarly, an expression like $f(x) \equiv x^2$ states that $f(x)$ is (*defined* to be) an abbreviation of x^2. Functions are a systematic way to specify synonyms for a class of objects. Related to the intensional/extensional example above, we'd want to say that 'human' and 'homo sapien' are synonymous and thus intensionally equal.

In MLTT, we also call definitional equality *judgmental equality*. A judgment can be viewed as a meta-logical claim. The proof of a judgment for MLTT is not within the logic, but external to it. For instance, when we want to say that a certain syntactic object is a type of the theory, we declare

$$A \ \mathtt{Type}$$

And this is *not* a proposition that can have terms as proofs. So too is the definitional equality of terms meta-logical in the sense that it is not a proposition of the language whose proofs are terms of that type. It would therefore be wrong to claim $a:b \equiv b'$. Indeed, any membership relation is a judgment, not a proposition. That is, the statement $a:A$ cannot have a proof term as it is not a proposition in our logic; it is a *judgment*. The distinction, although important, is not central and so I refer the reader, for a more thorough analysis, to [155, ch. 1].

On the other hand, propositional equality, when it holds, is an internally provable fact. In our type theory, this means it's a proposition that admits of proof, a type that admits of terms. The expression $2 + 2 = 4$ would be an instance of propositional equality. This is because we have to prove it, albeit easily; namely we couldn't say that 4 was *defined* to be $2 + 2$ merely because they 'refer to the same object'. Perhaps a clearer example is that we don't want to define 4 as the number of finite groups of order 63 up to isomorphism simply because they are the 'same number'. We define the (propositional) identity type $\mathtt{Id}_A(x, y)$ for a type A and terms $x, y:A$ to express the proposition that $x = y$ in type A. When the type

needs to be specified more clearly, we may write $x =_A y$.

A simple way to separate the two sorts of equalities is as follows:

Definitional The equality holds intensionally and is used to keep track of abbreviatory definitions, change in bound variables, and the principle of substituting defined terms for their definitions. These are also helpful constraints or stipulations for rules.

Propositional The equality is not simply true by definition; it needs a proof in the logic, no matter how simple.

An important feature of this specification of types should be noted: the definitional identity relation comes along with the type. In Martin-Löf's [91], he indicates that only definitional equality needs to be specified. Nordström, Petersson, and Smith reinforce this by saying that "A type is explained by saying what an object the type is and what it means for two objects of the type to be identical" [102, p. 7]. But since they use the notation $a = b \in A$ (this will correspond to our notation $a \equiv_A b$) to reflect the identity they mention, we recognize that it is in fact *definitional* equality. In the next section, I want to make the further claim that we also require identity types — propositional identity — defined for every type. Thus we'd have both definitional and propositional identity defined for each type.[10]

2.2.2 Identity Types

Now we get to specify the behavior of the identity types. We will also derive the symmetry, transitivity, and the trivial case of reflexivity from the rules given. This will serve as a basis on which our categorical interpretation will be explained in Chapter 6. Since identity is the focus of this paper, this section turns out to be very important.

The stipulation of identity types can have many formulations. Either we can view it as just another dependent type or we can view it as a necessary structure that needs to be included on any type. I take the latter of these views.[11] In much the same way that specifying a type requires a specification of definitional equality, I claim it also needs a propositional identity definition. That is, for any type A — including identity types themselves — we get 'for free' the identity

[10] I ignore here questions of size on types. We could relativize to universes and make precise what I mean by "every type" and so on. As this is rather orthogonal to my project, I will only suggest the concerned reader add the correct qualifications as we go.

[11] One is reminded of Quine's slogan "No entity without identity" [118, p. 20] and perhaps HoTT can be viewed as endorsing this in a modest way. We don't, of course, import Quine's general ontology nor do we need to deny it.

type on A. The motivation for this is similar with the motivation for requiring definitional equality for any type. The objects must have a criterion of (both types of) identity set out at the beginning. The definitional equality gives us a criterion for intensional identity. The identity type gives us a criterion for the concept of 'provably indistinct'. This notion will be expanded on as we go.

Consistent with the reading that every type needs an identity type, we have the type formation rule:

Identity Formation

$$\frac{A\ \mathtt{Type} \qquad a : A \qquad b : A}{\mathtt{Id_A}(a, b)\ \mathtt{Type}}$$

This natural deduction style of presenting type theory rules allows us to read this as follows: "Given a type A and two terms of that type, we have another type that corresponds to the proposition that those terms are provably equal". Note that this allows the nesting of identity types. Namely, we have the type $\mathtt{Id}_{\mathtt{Id_A(a,b)}}(p, q)$ for appropriate terms $a : A, b : A, p : \mathtt{Id_A}(a, b)$, and $q : \mathtt{Id_A}(a, b)$ etc. This provides a rich structure on identity types. Identifications between terms, identifications between those identifications, and so on. I fully admit that the intuition behind the nature of this rich structure of types is homotopical or more recently even *cubical* [17]. But I need not use this intuition for the formal proofs and features of the identity type in MLTT.

Among the inhabitants of certain identity types are the canonical self-equalities:

Identity Introduction

$$\frac{a : A}{\mathtt{refl_a} : \mathtt{Id_A}(a, a)}$$

This ensures that there's always a proof $\mathtt{refl_a}$ of the proposition $a = a$ for a term $a : A$. In other words, every term of every type is provably self-identical. This alone implies that propositional equality is reflexive. Indeed, that is why the special proof "$\mathtt{refl_a}$" is named as it is: to unconditionally assert that propositional identity is reflexive. Similar to all types in MLTT, simply saying that reflexivity is the only introduction rule for identity types is not enough to claim that each identity type is comprised *only* of reflexivity proofs. We'll also call the reflexivity proofs "trivial self-identifications", for convenience.

The identity elimination and computation rules are a bit more complicated. Informally, we can say that the elimination rule tells us that we only need to check the trivial self-identifications in order to claim something about all pairs of identified terms.

Identity Elimination (Path Induction)[12]:

$$\frac{x{:}A, y{:}A, p{:}\mathtt{Id}_{\mathtt{A}}(x,y) \vdash C(x,y,p)\ \mathtt{Type} \qquad x{:}A \vdash t(x){:}C(x,x,\mathtt{refl}_{\mathtt{x}})}{x{:}A, y{:}A, p{:}\mathtt{Id}_{\mathtt{A}}(x,y) \vdash J(t;x,y,p){:}C(x,y,p)}$$

This rule is presented in a sequent style to include the necessary variable contexts, but not with the general contexts. The first premise (top left) simply says that we have a type C that may or may not depend on two variables of type A and an identification between them. The second premise (top right) states that we have a term of C that only depends on one variable x and the trivial self-identification of x with itself. The rule allows us then to construct a special term J that takes all the information in the premises (i.e. the type C and the term $t : C$) to prove the proposition C for general $x, y{:}A$ and $p{:}\mathtt{Id}_{\mathtt{A}}(x,y)$. In essence, this tells us that we only need to establish there is a term of the dependent type C in the trivial case in order to get a term for any case. This is *Path Induction*. The homotopical interpretation of this rule gives it its name. If we view identifications (terms of an identity type) as paths in a topological space, then Path Induction allows us to 'induct' from the self-loop (which is the interpretation of `refl`) to all loops.

Now we'll go right into using this rule to get a feel for how it works. Specifically, I'll prove the usual symmetry and transitivity of (propositional) equality using Path Induction.

First we prove symmetry of propositional identity:

Theorem 2.1. *For a type A and terms $a, b{:}A$, if $p{:}\mathtt{Id}_{\mathtt{A}}(a,b)$, then we can construct a term $q{:}\mathtt{Id}_{\mathtt{A}}(b,a)$.*

Proof. For our dependent type C, which we call the *motive*, we choose $\mathtt{Id}_{\mathtt{A}}(b,a)$, which clearly only relies on $a, b{:}A$. Thus, using Identity Elimination, we get

$$\frac{a{:}A, b{:}A, p{:}\mathtt{Id}_{\mathtt{A}}(a,b) \vdash \mathtt{Id}_{\mathtt{A}}(b,a)\ \mathtt{Type} \qquad a{:}A \vdash \mathtt{refl}_{\mathtt{a}}{:}\mathtt{Id}_{\mathtt{A}}(a,a)}{a{:}A, b{:}A, p{:}\mathtt{Id}_{\mathtt{A}}(a,b) \vdash J(\mathtt{refl}_{\mathtt{a}};a,b,p){:}\mathtt{Id}_{\mathtt{A}}(b,a)}$$

Thus we have a term of type $\mathtt{Id}_{\mathtt{A}}(b,a)$ specified by J, giving us our q. □

The way to reason with Identity Elimination is to focus on finding an appropriate C that would land us near our goal. In the above case, we chose $\mathtt{Id}_{\mathtt{A}}(b,a)$ since that was exactly our goal. Now we prove the slightly harder case of transitivity.

[12] The notation here might seem complicated. This is still a natural deduction style, but with sequents that include their necessary variable contexts. Dependent types are indicated by the parameters they take in parentheses. For instance, $J(t;x,y,p)$ is term that is functional with four arguments. The semi-colon only indicates a conceptual (i.e. not needed) separation between the first and last three arguments.

Theorem 2.2. *For a type A and terms $a, b, c : A$, if $p : \mathtt{Id}_A(a, b)$ and $p' : \mathtt{Id}_A(b, c)$, then we can construct a term $q : \mathtt{Id}_A(a, c)$.*

Proof. If we try to use $\mathtt{Id}_A(a, c)$ as our C, it turns out to be rather difficult to come up with the appropriate contexts for the premises. So instead we choose a function type $\mathtt{Id}_A(y, c) \to \mathtt{Id}_A(x, c)$ which is well-defined when $x, y : A$. Note that c is fixed in this construction. This means that we have

$$a : A, b : A, p : \mathtt{Id}_A(a, b) \vdash \mathtt{Id}_A(b, c) \to \mathtt{Id}_A(a, c)\ \mathtt{Type}$$

the first premise of our rule. Further, if $a : A$, we have the trivial function $\mathtt{Id}_A(a, c) \to \mathtt{Id}_A(a, c)$ which is inhabited by the identity map id. This gives us

$$a : A \vdash id : \mathtt{Id}_A(a, c) \to \mathtt{Id}_A(a, c)$$

the second premise of our rule. Thus, by identity elimination, we have

$$\frac{a : A, b : A, p : \mathtt{Id}_A(a, b) \vdash \mathtt{Id}_A(b, c) \to \mathtt{Id}_A(a, c)\ \mathtt{Type} \quad a : A \vdash id : \mathtt{Id}_A(a, c) \to \mathtt{Id}_A(a, c)}{a : A, b : A, p : \mathtt{Id}_A(a, b) \vdash J(id; a, b, p) : \mathtt{Id}_A(b, c) \to \mathtt{Id}_A(a, c)}$$

giving us a term of type $\mathtt{Id}_A(b, c) \to \mathtt{Id}_A(a, c)$. But we have a term $p' : \mathtt{Id}_A(b, c)$ so this function gives us the needed proof (term) q of $\mathtt{Id}_A(a, c)$, which proves transitivity. □

Understanding Identity Elimination is crucial to the project here so I will elaborate on what I take to be the important features. As in the case of symmetry and transitivity, the particular C didn't actually depend on the identification $p : \mathtt{Id}_A(x, y)$. This case illustrates why the rule is called Path Induction. We have a type C that can apply to any two objects of A and we want to know whether it holds in general. The second premise of Path Induction only requires us to say that we can prove the proposition C holds for the pair (x, x). This means that C is a reflexive relation on A. Whether C itself relies on the identification $p : \mathtt{Id}_A(x, y)$ or not, the conclusion of Path Induction tells us that showing C to be well-defined and reflexive is sufficient to conclude that C holds for all *provably identical* terms of A. This allows us an intuition on how Path Induction works.

How to think about Path Induction

We can prove something (the motive) about all equal pairs of terms by only considering the trivial self-identifications.

However, Path Induction does not stand on its own. We also have a *computation rule* for identity types that specifies the behavior of the canonical term J.

Identity Computation:

$$J(t; x, x, \mathtt{refl}_x) \equiv t$$

That is, if we give J the trivial case, we get nothing more than the second premise back.

And now we prove the final important feature of the identity types that will concern us here. We have not distinguished levels of our type theory to indicate when a type is a `Type`, `Set`, `Prop`, etc. But we are going to prove something about 'predicates' in the type theory. We can, for our purposes, view predicates as dependent families into `Type`. So a predicate $P(x)$ with the indicated free variable will be written as $x{:}A \vdash P(x)\ \mathtt{Type}$. The next theorem is a type theoretic version of Leibniz's Law of the indiscernibility of identicals. Put logically, it will say that if $P(a)$ is true and $a = b$, then $P(b)$ is also true, i.e. provable. Often, this result is considered to be the Id-Elim rule, which is interesting since Path Induction (Id-Elim in HoTT) proves it.

Theorem 2.3 (Leibniz's Law)**.** *If $a, b : A$, $p : \mathtt{Id}_A(a, b)$ and for some predicate $x{:}A \vdash P(x)\ \mathtt{Type}$ over A we have $q{:}P(a)$, then there is a term of type $P(b)$.*

Proof. We use identity elimination with the motive $P(x) \to P(y)$ since we want a way to *transport* proofs of $P(x)$ to proofs of $P(y)$ for equal terms.

$$\frac{x{:}A, y{:}A, p{:}\mathtt{Id}_A(x,y) \vdash P(x) \to P(y)\,\mathtt{Type} \qquad x{:}A \vdash id{:}P(x) \to P(x)}{x{:}A, y{:}A, p{:}\mathtt{Id}_A(x,y) \vdash J(id; x, y, p){:}P(x) \to P(y)}$$

Clearly the premises hold and so we have an instantiation

$$J(id; a, b, p){:}P(a) \to P(b).$$

And therefore applying this function to $q{:}P(a)$ gives us a proof of $P(b)$ as desired. □

So far, we've proven that the identity type is reflexive, symmetric, transitive, and substitutive (Leibniz's Law). We will do the same for a categorical characterization of identity given later. The reader should note how the above features were proven as their proofs will be largely mimicked in the later discussion, giving force to the argument that they are the same concept. Once I've argued that the identity type and the independent notion of identity arrived at later are the same in the relevant respects, their justification will also be the same. Thus, the mathematically motivated and informed characterization of identity will be justified on fully mathematical grounds which will give Path Induction similar justification. In a sense, the justification problem is solved by reducing the MLTT identity rules

to a general notion of identity. Thus, for MLTT and HoTT to serve as foundations, they need to represent the general *mathematical* notion of identity which is demonstrated in this paper.

I would also point out that showing the motive of Identity Elimination to be reflexive is indeed the crucial part as explained categorically in Chapter 6:

Identity is that relation that implies all and only reflexive relations.

This characterization of identity will have to wait a bit, however, as we need to justify the mechanism by which it is constructed. So we first turn to situating our project in the context of Ladyman and Presnell's recent work and then go on to lay out what it would take to give a characterization of identity. Near the end of the paper, in Chapter 6, we will elaborate on the above characterization and prove it is equivalent to the identity rules of MLTT.

3 Ladyman, Presnell, and Foundations

As stated earlier, my project is motivated by recent work by James Ladyman and Stuart Presnell in justifying HoTT as a foundations. Although I take issue with the constraints they put on such a program, I am altogether sympathetic with their project and will offer what I take to be an alternative account. Foundational concerns loom large in the philosophy of mathematics and here is where we first engage the philosophical literature with respect to the justification and the foundational status of a particular program. The focal point of the foundational debates, I believe, is on what it means for a system to be properly foundational. Some require the putative foundation to exhibit pre-theoretic ties in some ways. The idea being that a foundation of mathematics cannot in turn rely on mathematics and so it must be relevantly supported by external systems and concepts. Ladyman and Presnell require such an external tie: pre-mathematical justification.

I will take a different tack and avoid such an external constraint on foundations. There are two main ways to do this, that I see. The first is to ignore questions of foundations in any of the epistemic, logical, and conceptual senses. This would surely avoid, by dismissing them outright, constraints on foundations since foundations would be neither desirable nor relevant to the anti-foundationalist. Although I think there is some merit in such a position, I will take yet another approach. I will be giving a *mathematical* justification and explanation of HoTT (at least one part of it) as a putative foundations. The presumed circularity of this position will ring loud in the ears of those who desire external justification of a foundation; indeed the motivation for such an external justification was precisely to avoid this circularity. I will try to deal with this worry below when discussing in detail the criterion of justification.

The plan for this section includes first characterizing the justification Ladyman and Presnell give for Path Induction. Here I find their justification plausible yet I also find a few important features I think they may have missed. Then I will outline the underlying philosophical program they use when discussing foundations of mathematics. It is here that I diverge more strongly from their position; I bring

a different set of foundational constraints to the problem of justification. My goal, however is not negative. I aim to provide a different account of the justification of HoTT's rules of inference that follows from a different set of success criteria for a foundations of mathematics.

It is worth mentioning that the two papers I take to indicate their position, [70] and [69], are works in progress. I recognize that it would be unfair of me to pretend they were not so. It does not matter much what the status of these papers is, for me, since my project is largely sympathetic and not antagonistic; the worry that I mischaracterize their argument is small given that I do not really aim to refute it. They do provide the motivating challenge that this project is an answer to, but I feel compelled to say at the outset that I take their work only as a springboard for my project and not something I will spend much time arguing against.

3.1 The Argument

Path induction is the term used for the elimination rule for identity in HoTT. As we've seen, one consequence of its formulation is that in order to make any statement about identifications of two terms (or those identical terms themselves), we need only check the trivial self-identification of one of those terms. In other words, it is sufficient to check the trivial self-identifications to make statements about all terms of the identity type. This is a sort of induction, where we might naïvely imagine that we only need to check the base case to cover the whole type. The word "path" comes from the homotopical interpretation that views identifications as paths between points or, importantly, higher dimensions of 'path' between higher dimensional objects like lines, squares, cubes, and so on.[1] So identifications between points are the paths and the identifications between paths are the simplices or squares and the identifications between squares are cubes, etc. At first, this interpretation seems irreducibly mathematical.

Ladyman and Presnell use two basic principles to justify Path Induction without homotopy or path objects: the *uniqueness principle for identity types* and *substitution salva veritate*. For the first, they take examples like that of the product of two types to motivate. Not only are pairs the canonical terms of the product, we can express and prove that every term in the product is (evaluated to) such a pair. This is the *uniqueness principle* for products. To restate this, the introduction rules says that pairs of terms inhabit the product type. In addition, the uniqueness principle for products states that *every* term of the product type is a pair. The case is a little more subtle for a uniqueness principle of identity types.

[1] I use the word "identifications" as shorthand for the equivalent notions of 'a proof of an equality' and 'a term of an identity type'.

Recall the introduction rule for terms of identity types

$$\text{Identity Introduction:} \qquad \frac{a : A}{\mathtt{refl_a} : \mathtt{Id_A}(a, a)} \ .$$

The problem for a uniqueness principle is that the type former for identity types,

$$\text{Identity Formation:} \qquad \frac{A\ \mathtt{Type} \qquad a : A \qquad b : A}{\mathtt{Id_A}(a, b)\ \mathtt{Type}}$$

gives us types that mention two terms of type A. Thus to say that every term of type $\mathtt{Id_A}(a, b)$ is $\mathtt{refl_a}$ doesn't make sense; they're of different types. What Ladyman and Presnell stipulate then is that we reform the uniqueness principle to state that $\mathtt{refl_a}$ gives us "all identifications *up to identity in the appropriate type*" [70, p. 19, original emphasis].

To make the uniqueness principle precise, we need to nail down in which type we're identifying $\mathtt{refl}$ with something else. We want to say something like "every $p : \mathtt{Id_A}(a, b)$ is identical to $\mathtt{refl_x}$ for some x". But $\mathtt{refl_x}$ and p are obviously not in the same type, so we have to find a type in which we can say something about their propositional equality. Define the dependent sum

$$\mathtt{E}_a :\equiv \sum_{x:A} \mathtt{Id_A}(a, x).$$

In the literature, this is a singleton type since it holds one of the arguments fixed. But it is equivalent to the general identity type as shown simply in [29] where our $\mathtt{E}_a$ corresponds to Coqand's $\mathtt{singl}(A, x)$ type. Indeed, this was the identity type first considered for MLTT by Paulin-Mohring [105].

For us, $\mathtt{E}_a$ is the type that consists of pairs (b, p) where $b : A$ and $p : \mathtt{Id_A}(a, b)$. In our interpretation of terms as proofs of propositions, this is the type of proofs that a term b is identical to a in type A. As a dependent type, when we change b, clearly the way you show b is identical to a may also change. Naturally, $(a, \mathtt{refl_a}) : \mathtt{E}_a$ since $\mathtt{refl_a} : \mathtt{Id_A}(a, a)$. We then express the uniqueness principle for identity types in $\mathtt{E}_a$ as

$$\prod_{(b,p):\mathtt{E}_a} \mathtt{Id}_{\mathtt{E}_a}((a, \mathtt{refl_a}), (b, p)) \tag{3.1}$$

which says that every pair $(b, p) : \mathtt{E}_a$ is provably identical to the pair $(a, \mathtt{refl_a}) : \mathtt{E}_a$. We stipulate that this type is inhabited, which means that we have a proof of the proposition that these pairs are equal.

Before moving on to the second principle, I will mention that their terminology "uniqueness principle of identity types" might suggest something like the "uniqueness of identity proofs" (UIP) which states that there is only one proof of equality.

This would indeed be most unhelpful in justifying Path Induction. If we accept that there is only one reason (i.e. proof) that two terms can be identical, then we are no longer even in Homotopy Type Theory, let alone with an interesting Path Induction rule. The reader should make sure to note that uniqueness of the terms in the identity type as used by Ladyman and Presnell does not imply the extensionality of the type theory, as UIP would. In effect, the *uniqueness of identity types* is a philosophical way of stating the *contractability* of the identity types. This is a homotopical notion, which further confuses me as to why the authors feel comfortable asserting that *uniqueness of identity types* is a pre-mathematical premise. Indeed, I can't think of any pre-mathematical reason why terms of identity types should be unique *up to propositional equality* in a further type — i.e. contractible. It might reasonably be argued that the pre-mathematical notion of identity cannot justify multiple proofs of identity, so the path interpretation of identity types would be impossible to justify in their framework. So if they wanted to assert UIP, they'd remove any possibility of justifying Path Induction, but the principle of *uniqueness of identity types* on the other hand seems to be based on a strictly homotopical intuition. I cannot see how their project thus stated can incorporate this principle consistently. Another way to put this is to consider how we are to pre-mathematically *distinguish* between UIP and the uniqueness principle for identity types. Are our intuitions fine-grained enough for this task? I would say they aren't, but this would require detailing what intuitions are capable of. Thus, this is only a *prima facie* objection that I shall leave here.

The second principle Ladyman and Presnell use is the principle of substitution *salva veritate*. This is the belief that identical objects should be inter-substitutable in propositions while preserving truth values. This is called *transport* in the HoTT literature and is widely used. Briefly, we can state is as the existence of a function

$$\mathtt{Id}_{\mathtt{A}}(a,b) \times P(b) \to P(a)$$

for predicate P over A and terms $a, b : A$. This says that if we can prove $a = b$ in type A and also that $P(b)$, then we can prove $P(a)$. This works for any type so we can replace A with $\mathtt{E}_a$:

$$\mathtt{Id}_{\mathtt{E}_a}((a, \mathtt{refl}_{\mathtt{a}}), (b,p)) \times P(a, \mathtt{refl}_{\mathtt{a}}) \to P(b,p) \tag{3.2}$$

The uniqueness principle (3.1) states that $\mathtt{Id}_{\mathtt{E}_a}((a, \mathtt{refl}_{\mathtt{a}}), (b,p))$ is inhabited, so (3.2) gives us a function $P(a, \mathtt{refl}_{\mathtt{a}}) \to P(b,p)$, which is exactly the induction from $\mathtt{refl}_{\mathtt{a}}$ to all identifications of $a : A$ with any $b : A$. Thus, Ladyman and Presnell claim they have set Path Induction on pre-mathematical grounds and HoTT is now that much closer to being an autonomous foundation for mathematics. This of course requires that their two principles, that of uniqueness for (identity) types

and substitution *salva veritate*, be pre-mathematically justified. They give no indication why the uniqueness principle for types is justified on pre-mathematical notions, but they do indicate that substitution *salva veritate* is part of the intuitive notion of identity. Instead of arguing against the pre-mathematical nature of these two principles, I will be arguing that the very requirement that they must be pre-mathematically justified ought to be dropped.

Mathematically speaking, the literature has recognized that Path Induction is equivalent to the combination of *transport* and *contractibility* [27], which are the mathematical ways of stating the two principles used by Ladyman and Presnell. Although the use of this fact is different in the Ladyman and Presnell project than it is in other contexts, the motivation is quite clear: This breakdown of Path Induction is a useful conceptual tool in analyzing identity types. I will return to this point in Chapter 6.

3.2 The Criterion for Justification

We now come to the central question in foundations: What counts as a foundation? The properties of each presumed foundation are certainly not agreed upon among philosophers. But I'd argue that the most pervasive difference among the varied accounts is their criterion of success. When Feferman argues against the proper foundational nature of category theory in [40], it is based on a requirement that category theory satisfactorily captures a collection of desirable mathematical objects without reference to other foundations. When Linnebo and Pettigrew take a similar position in [81], they ask whether category theory is sufficiently autonomous (conceptually, logically, and with respect to justification) from other foundations and areas of inquiry. Unfavorable reactions and sympathetic accounts[2] all exhibit a basic difference: their criterion for success.

In their forthcoming paper on the prospects of HoTT as an 'autonomous' foundation for mathematics, Ladyman and Presnell give some explicit criteria that they think any autonomous foundation must satisfy. I reproduce the criteria from [69, p. 4]:

1. A single framework in which to cast some or perhaps all mathematical theories. This framework involves a mathematical language and theory.

2. A semantics in the sense of an account of how the theoretical terms of (1) are to be understood, how concepts are related (for example, higher level concepts being composed of lower level ones, all concepts being defined in terms of some basic concepts, and so on), and an account of how the rules that are used to manipulate the concepts are to be understood and justified.

[2] See [9], [56], [5], [96].

3. A metaphysics that may include an ontology and such notions as mind independence or mental construction.

4. An epistemology in the sense of an account of how we are able to know the truths (if any) of mathematics, given (2) and (3).

5. A methodology for mathematical practice based on some or all of the above.

We will restrict attention to (2): semantics since our focus is about rules of inference. Specifically, they claim that a foundations for mathematics requires "an account of how the rules that are used to manipulate the concepts are to be understood and justified". So we need to justify our rules of inference, which I will outline my preferred method in the next chapter. But we have the further requirement that when we are justifying our foundations we are "providing a grounding or basis in pre-mathematical ideas" [69, p. 3]. It is this condition on our justification that seems unwarranted. The justification of foundational concepts in terms of pre-mathematical notions is not necessary and there are several criticisms to make.

First, we notice that in the case of type theory, the pre-mathematical requirement amounts to saying the entire type theory must be a formalization of pre-mathematical notions. This is because formal type theory, under this criterion, would need to be justified completely by pre-mathematical concepts. I cannot see how this would require any less than a reduction of the formal language to an informal, or pre-mathematical, one. Why this should be required is not quite clear and indeed seems counter to the purpose of formalization. It seems a heavy burden on imprecise and ambiguous intuitions and notions to carry the full formal framework of HoTT. Although I do think plenty of type theory has its germ in pre-mathematical notions, I do not think this is any more than a simple truism: our theoretical concepts are influenced by our pre-theoretical concepts. But the requirement that *all* of our theoretical concepts be justified completely on pre-theoretical concepts is much too strong.

The pre-mathematical requirement forces us to say that every new concept we find in an autonomously founded mathematics is justified on pre-mathematical grounds. I assume that under this picture, justification is somewhat transitive so that the justification of the foundation of mathematics gives us some justification of the higher mathematics. For example, a justification of set theory as a foundation will give us justification of the mathematics built upon set theory. But this is to presume that our pre-mathematical concepts are robust enough to justify *all* of mathematics. It's unclear whether we should view this as overestimating the ability of pre-mathematical ideas to support mathematics, or restricting mathematics to an extreme degree to those concepts with a pre-mathematical justification. In

either case, I think we make a mistake in bringing them to the same level or reducing mathematics to the pre-mathematical.

Set theory is taken to be a likely candidate for the sort of foundation Ladyman and Presnell are looking for. We can hold that set theory — we have a full menu of set theories now — formalizes the pre-mathematical notions of collection and membership. I do accept that the concept of 'set' seems to be a precise formulation of the pre-mathematical notion of collection. And I also accept that the concept of '$\in$ membership' is a precise formulation of the pre-mathematical notion of element or member. But I do not think this justification runs up the tree of set theoretic knowledge to things like large cardinals. Under this reasoning, we would even treat axioms like that of infinity with suspicion. So although our pre-mathematical notions get us in at the ground floor, they don't —and I argue that they shouldn't be expected to — justify all of set theory.

Thus even in the paradigm case of set theory, the pre-mathematical requirement is too strong. If Ladyman and Presnell think that set theory does provide an autonomous foundation for mathematics, then they must accept something like the following tiered approach: pre-mathematical notions justify primitive notions, but mathematics proper lets us continue. What is perhaps striking is that Ladyman and Presnell believe lambda terms to be pre-mathematically justified by the notion of substitution like that for pronouns in natural language.[3] So it seems that we can get in at the ground floor in type theory as well. Indeed, the authors agree with Mayberry when he says

> [T]he primitive concepts of mathematics are those in terms of which all other mathematical concepts are ultimately defined, but which themselves are grasped directly, if grasped at all, without the mediation of a definition. [94, p. 18]

The focus is that the primitives must be grasped pre-mathematically. The authors give no indication why the identity type is such a primitive in need of 'mediation-less' grasping. The line between primitive and defined is a slippery one to use when giving a requirement for foundations. We might for the time, however, allow that rules of inference are sufficiently primitive.

The authors do not claim that any of this is a death sentence for HoTT's foundational position. They want to explore in particular whether HoTT can be an *autonomous* foundation. Their use of "autonomous" is similar enough to invite comparison to [81]. And the pre-mathematical requirement is leveraged to get that

[3] The wording is as follows on [69, p. 9, note 16]: "While lambda calculus is a sophisticated domain of study, the basic notion of substitution in an expression is a simple pre-mathematical one, familiar to anyone who, for example, is able to use pronouns in natural language. We therefore do not consider this to be an obstruction to the autonomous status of HoTT."

autonomy. So the foundational status of HoTT does not hang in the balance, but the methodology for showing it to be an autonomous one does. I argue that we can do better than to require a pre-mathematical justification for every mathematical concept and rule. But I will restrict my attention to the identity type rules, as they do in [70].

Ladyman and Presnell establish what they take to be a pre-mathematical *presentation* of Path Induction. This presentation as sketched above (§3.1) uses type theoretic and mathematical notation and concepts. They do avoid use of homotopy, which was their goal. But to view homotopy as somehow more sophisticated or too mathematical to be used in justification is unargued for. What criterion of simplicity or sophistication of mathematical concepts are they using? Why is it appropriate to use

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

in their presentation but not path homotopy? Indeed, Ladyman and Presnell take issue with the *presentation* of Path Induction. So we are allowed to judge their presentation on whether it's pre-mathematical or not. It's hard to argue that their reasoning, given the uniqueness principle for identity types, is presented in a pre-mathematical way. This brings us to the second objection.

The line between the pre-mathematical and mathematical is vague. Although not useless, it does beg analysis if we are to rest our foundational concerns on the satisfaction of such a pre-mathematical requirement. The notion of pre-mathematical gets used in several ways. The way that I think is most problematic is when it stands for the capacity of an individual to *understand* a concept before understanding others, as in Mayberry. This was emphasized in [81] as 'conceptual autonomy'. We need to understand homotopy before we can understand Path Induction properly. So, it is argued, Path Induction cannot, as is, contribute to an autonomous foundation, for it is not itself autonomous. But this is to assume a robust, if not defined, categorization of the mathematically sophisticated and the relatively primitive. We are not talking about actual logical primitives, but in a sense, the 'ground floor' for the foundations under considerations.

As no one has yet to my knowledge presented such a robust categorization (especially based on pre-mathematical ideas), we are left wondering how we are to distinguish between the sufficiently basic from the sophisticated. It strikes me as more likely that many of the distinctions we make about mathematical sophistication are largely based on cognitive and historical facts. It is true that most people learn homotopy after they learn about equality, so defining equality in terms of homotopy feels backwards, in some sense. It is also true that homotopy came long after notions of mathematical identity were on the table. And this might make it seem like homotopy is necessarily more "sophisticated" then identity. But

it bears noticing that identity is a much more slippery notion in general than homotopy, which is simply defined. If this is a mark of the pre-mathematical nature of identity, then we are back to the first objection of whether our pre-mathematical intuitions are robust enough. That is, if pre-mathematical notions like identity are more ambiguous or have underdetermined semantics, then why are we to rely on them for justifying all of mathematics?

Ladyman and Presnell, in their papers, do not give a positive argument why homotopy itself is not pre-mathematically justified. We might think of the intuitive notion underlying homotopy to be something like the path identity we use in our spatial reasoning. If you tell someone to walk down this path to the end, they will allow themselves slight deviations, due to the way we walk, from that path. There seems to be some intuitive notion of path identity operating here that allows the identification of spatially distinct paths. In correspondence, the authors point out that this is plausibly due to a pre-mathematical notion of distance or approximate distance rather than path identities. They also point out that we might intuitively think that going around a tree (in a loop) would be roughly the same path where it is not homotopically the same. Both of these points are well-taken and I think it shows that homotopy is not completely captured by this brief example of a path one walks upon.

I would similarly argue that pronoun substitution does not completely capture lambda terms. In a sense, we're agreed that the pre-mathematical doesn't capture *everything* about the mathematical concept. Pronoun substitution doesn't — at least for my intuitions — capture the entire computational baggage associated with the lambda calculus. It seems then that we need to capture the 'essential' parts of the mathematical concepts; we need not require that the intuitive notion of substitution make all the properties of the lambda calculus obvious. So we come to a problem of differentiating the essential or important components of a concepts from the those that need not be captured by intuitions and pre-mathematical notions.

I leave the objections to the criterion for foundations Ladyman and Presnell put forward here. Since my project is mostly positive, I take the above objections to foreshadow the alternative that will be given below. This criterion provides a useful foil to make clear what sort of criterion I want to defend. The feature that will be dealt with most below will be the very requirement that a foundation of mathematics requires a justification and explanation in terms of things *not mathematical*. Although common, this program of an 'externalist' justification will not be endorsed in my project. I think we are better served under a weakly holistic view of mathematics that allows *mathematical* justification for the foundations of mathematics. I will make some points with respect to the apparent circularity in my discussion of Michael Dummett's work *The Logical Basis of Metaphysics* [34]

in Section 4.2.

The general scheme for justifying Path Induction is to reduce it to a feature of a general categorical logic in the major theorem of this paper. By proving that the categorical conception of identity implies Path Induction, we show that Path Induction is justified. If we accept category theory as part of mathematical practice, the theorem shows that Path Induction is consistent with mathematical practice. If we accept category theory as a foundation of mathematics, then the theorem shows that Path Induction is foundationally justified. If we are concerned that Path Induction is of a different nature than other mathematical conceptions of identity, the theorem should assuage those concerns. But first, we will consider the prospect of justifying rules of inference in general. Then, after remarks about the meaning theory that is determined by justified rules of inference, I will prove the theorem that justifies Path Induction.

4 Justifying Rules of Inference

4.1 History

The history of justifying rules of inference is a long and rich one. Justification itself relies on implicit or explicit rules of inference and it has been at the center of the philosophical enterpise from the beginning. The *reasons* someone can give for a claim are often supposed to *entitle* them to that claim. I believe this sort of norm-based view of inference rules is most appropriate in natural language, but I will also use it for the logical and mathematical case. A rule of inference can be expressed in formal frameworks as a sequent $\Gamma \vdash \Delta$ where both Γ and Δ are sets of formulae. Assuming for the moment that we have a language rich enough to capture any sentence we might want to express, we can read this sequent as stating a certain rule of the language game: If you establish (assert, know, believe, etc.) those sentences in Γ, then you are 'allowed' to establish (assert, know, believe, etc.) any formula of Δ. The most natural cases usually have Δ as a singleton $\{\phi\}$ so we can assert a sentence ϕ based on Γ.

But in studies of proof in mathematical logic we usually find it more convenient and general to use sequents themselves as the antecedent and consequent of rules of inference. We represent inferences with a line

$$\frac{\Gamma_1 \vdash \Delta_1 \qquad \Gamma_2 \vdash \Delta_2 \qquad \ldots \qquad \Gamma_n \vdash \Delta_n}{\Gamma \vdash \Delta}$$

And we say that we can establish the sequent below the line if we've established all the sequents above the line. This is for finite premises and we could extend this to the infinite case but will avoid doing so for our current project. We will also require the sequents to be single-succedent sequents;[1] this is where Δ_i is a singleton. An intuitive example is that of $\wedge$-introduction:

$$\frac{\Gamma_1 \vdash A \qquad \Gamma_2 \vdash B}{\Gamma_1, \Gamma_2 \vdash A \wedge B}$$

[1] The antecedent is to the left of $\vdash$ and the succeedent is to the right.

which we read as saying that if you can prove A with formula in Γ_1 and B with formula in Γ_2, then you are allowed to assert $A \wedge B$ with the sentences in Γ_1 and Γ_2.

Note the pattern based way rules of inference are stated. We are allowed to replace A and B with any sentences or propositions we like. But we also don't restrict the sort of formulas in the antecedent of the sequents so this rule is a schema for a potentially infinite number of inferences.

We use similar, if not so formal, reasoning in empirical matters as well. What are the conditions under which saying "that patch is red" is warranted? The answer certainly integrates quite a few implicit assumptions and if we could flesh these out, they'd go into our rule as antecedents. The way to think about the antecedents can vary based on context. The two ways I think are most appropriate are as follows:

1. The speaker (who aims to assert the conclusion) must make sure he or she is justified in asserting each sequent above the line before they're justified in asserting the sequent underneath the line.

2. The listener can infer from the speaker's assertion that the speaker must have justification for their antecedent committments.

Of course there isn't much different between these two perspectives. But it should be noted that a prominent reason why logic developed in this way was to make problem solving an *interpersonal* project. So it shouldn't surprise us that the speaker's commitments and the listener's entitlements are the same. There is some messiness in (2) since there may be many possible configurations that justify the speaker's assertion. We might ask, is the listener to assume a *canonical* way of justifying the speaker's assertion?

The way this is formulated here usually suggests a foundational picture. After all, if the justification of the speaker depends on their justification of the antecedents and those are justified by futher antecedents and so on, to avoid both circularity and infinite regresses, there needs to be something like axioms or first principles. I do not take this 'suggestion' and indeed find it to be overreaching on the intuition given by the sequent style reasoning. Although we do have axioms and unjustified primitives, I don't think we need to have such stringent requirements on our justification story. I will not argue for a different program of justification, but something more coherentist and pragmatic would be more in line with my view. I take the foundational view as a special case, or an extension of the norms of justification, of mine. It is not so much wrong as it is too strong. In fact, I will not be supposing a foundational story in the analysis below since it is not required for justification to make sense. I would certainly like to ward against any reductionist justifications of mathematics or type theory. I think the general

set of objections to empiricist reductionist programs to be largely generalizable to anti-reductionist arguments in mathematics. But I shall not take up this thread here, although hints of it were given in the previous chapter.

Before diving in to the history, I should note that similar to Dummett in [34], I treat axioms or 'logical laws' as akin with rules of inference. In our sequent presentation, an axiom φ can be expressed as the rule

$$\frac{}{\vdash \varphi}$$

with or without an antecedent of the sequent. In Gentzen presentations we can prefer the sequent axiom rule

$$\frac{}{\Gamma \vdash \Gamma} \cdot$$

Since we can make many, if not all, axioms or logical laws into rules of inference, the problem of justifying rules of inference includes the problem of justifying axioms. But it may seem like a very different project to many.

If we do not restrict ourselves to a particular domain of language, there is a worry that our justificatory story will be inherently circular. Not least among the philosophers who have posited such a worry is Wittgenstein [162]. We must, it seems, keep some rules of inference implicit if we are to avoid this circularity. But for mathematics, and especially type theory, we make the rules of inference explicit, for what should be clear reasons.[2] Lewis Carroll also points out in [25] how reasoning, even formally, prevents us from turning all rules of inferences into premises, axioms, or 'laws'. So rules of inference are ineliminable and of course central to logic. Below I'll give a brief and very incomplete history of the attempt to justify logical rules that will land us near an inferentialist account of the rules of inference.

Clearly, the role that logical inference rules are supposed to play will impact how we justify them. For the philosopher of science, there are particular standards that seem most attractive. From Peirce's leading principles, we have materially true premises and materially true conclusions related in an inductive or probabilistic way. But for an initial justification of logical rules, we turn to Carnap's work in [23]. Traditionally, rules of inference are judged by their semantic rather than syntactic features. We can view this in part as an artifact of the prevailing interest of rules of inference in the philosophy of science especially for induction. This is because we have an empirical world, and ideas about that world, that we feel should be represented validly in our inferences. Although my project is not directly related to the philosophy of science, we use the empiricist as a good starting point for the sort of considerations that inform what we think it takes to justify

[2] One of the reasons is precisely the formalization and mechanizability project that HoTT is so intertwined with.

a rule of inference. Carnap proposes reversing the traditional semantic order of explanation:

> Up to now, in constructing a language, the procedure has usually been, first to assign a meaning to the fundamental mathematico-logical symbols, and then to consider what sentences and inferences are seen to be logically correct in accordance with this meaning. Since the assignment of the meaning is expressed in words, and is, in consequence, inexact, no conclusion arrived at in this way can very well be otherwise than inexact and ambiguous. The connection will only become clear when approached from the opposite direction: let any postulates and any rules of inference be chosen arbitrarily; then this choice, whatever it may be, will determine what meaning is to be assigned to the fundamental logical symbols. By this method, also, the conflict between the divergent points of view on the problem of the foundations of mathematics disappears. For language, in its mathematical form, can be constructed according to the preferences of any one of the points of view represented; so that no question of justification arises at all, but only the question of the syntactical consequences to which one or other of the choices leads, including the question of non-contradiction. [23, p. xv]

Carnap's views developed significantly over his career and so we do not pretend to ascribe a fixed position here to Carnap the man. Indeed, Carnap's attention to semantics as well as pragmatics was such that the above view might not fit with his later views even in the same work. We will not be using 'pragmatic' below in the way he used it in [23, p. 454] where he says "Our considerations belong, strictly speaking, to a biological or psychological theory of language as a kind of human behavior, and especially as a kind of reaction to observations." For us, this is the 'language-entrance' move that we will not be discussing in detail. For now, we use Carnap's view as a jumping off point for our program of justifying rules of inference.

The classic response from Prior [116] showed why the rules can't be "chosen arbitrarily" as Carnap wished. Prior introduced a connective called *tonk* that has the introduction rule of disjunction and the right elimination rule of conjunction. If we denote tonk by $\star$, then the rules are

$$\frac{A}{A \star B}\,\star\mathrm{I} \quad \text{and} \quad \frac{A \star B}{B}\,\star\mathrm{E}$$

The succession of $\star$I followed by $\star$E would allow the derivation of $A \vdash B$ for all A and B: the runabout inference ticket. Each rule *by itself* is perfectly legitimate.

Prior's point is that their *combination* necessarily factors in into whether they are legitimate or workable rules of inference.

So not just *any* inference rule and associated connective would work, given certain norms of logic like the avoidance of inconsistencies. That is, there's more to be said than what Carnap has indicated. The received view that Carnap opposes in this quote we might call the semantic approach to rules of inference. On this view, we have an prior meaning in mind when we construct connectives. Carnap's view then would be called syntactic. Of course we could view the tonk objection as the demise of Carnap's syntactic approach. We might, in the end, need to reason semantically. This would block tonk, for it does not represent any previously accepted concept. Stevenson, for instance, in [145], advocates going back to the order of explanation that Carnap was opposing. But we need not abandon our project quite yet.

Belnap's response [11] in turn was to differentiate good connectives and rules from bad ones by considering *conservative extensions*. In brief, if $\mathfrak{L}$ is our language and rules and we add a connective with some rules to get $\mathfrak{L}^+$, $\mathfrak{L}^+$ is a conservative extension of $\mathfrak{L}$ if all the propositions stated in $\mathfrak{L}$ do not change truth value in $\mathfrak{L}^+$. This is a well established concept in logic now. It does block things like tonk. For A and B would be in $\mathfrak{L}$ and if $\mathfrak{L}$ was consistent $\mathfrak{L} \not\models A \to B$ for some A and B.[3] But when we add tonk, we get $\mathfrak{L}^+ \models A \to B$ for all A and B. Thus a non-theorem of the original language has become a theorem in the supplemented language. So tonk is out.

Although we can convince ourselves, perhaps, of the necessity of this conservative extension requirement, it isn't obviously sufficient for inference rules. The condition really only says that the new connective, with its rules, has to be relevantly independent of the initial logic. It is not allowed to interact with the connectives already there in any unsavory way. But is this enough? If it is, then the internal structure of the rules of inference is of little to no consequence. Belnap's suggestion would amount to giving mostly external reasons for accepting a rule. And this, I claim, is problematic. We want at least part of the inference rules to be self-justifying. This is in line with a more syntactic approach to logic. Syntax is internally justified, and not referentially or externally so, as semantics seems to be.

Prawitz pursues exactly this sort of self-justifying program. He wants the introduction rules to be self-justifying.[4] The elimination rules, in contrast, are justified only by reference to the introduction rules. The criterion, for Prawitz,

[3] Namely, a consistent system will not model $\top \to \bot$.

[4] The word "self-justifying" might put many readers on guard. I only mean that the introduction rules have no constraints to check. If we take an inferentialist perspective, we might say that this is not a case where 'asking for reasons' is appropriate. Relatedly, we might even say that the introduction rules are arbitrary, like Carnap wanted.

for the elimination rules is normalizability, or the ability to locally reduce and expand the appropriate formula in a derivation. The natural deduction calculus must be locally reducible with the new rule and connective. Clearly tonk ruins any hope of normalizing the derivations in the logic. Detour formulas would not be eliminable in general. For example, we can systematically show that occurrences of $\wedge$I followed by one of the $\wedge$E rules can be eliminated in favor of more direct reasoning:

$$\dfrac{\dfrac{\begin{array}{cc}\mathcal{D}_1 & \mathcal{D}_2 \\ A & B\end{array}}{A \wedge B}\,\wedge\text{I}}{A}\,\wedge\text{E1} \quad \text{reduces to} \quad \begin{array}{c}\mathcal{D}_1 \\ A\end{array}$$

But it should be clear that there is no hope of reducing similar detours of tonk:

$$\dfrac{\dfrac{\begin{array}{c}\mathcal{D}_1 \\ A\end{array}}{A \star B}\,\star\text{I}}{B}\,\star\text{E} \quad \text{reduces to} \quad ?$$

where in general we don't have a proof of B ready to hand.

We might be concerned that this too feels like an external requirement for rules, much like the conservative extension requirement. And although this is partially true, Prawitz only applies the normalization requirement on the elimination rules. The introduction rules can be included in our logic without any worry about conservative extension, for the only changes to the logic by adding introduction rules are those sentences with the new connective — so there are no changes in the original language. But the elimination rule must maintain normalizability with respect to the introduction rule. This may feel like a compromise between Belnap and Carnap. Whereas we allow arbitrary introduction rules, we restrict the elimination rules to be meta-mathematically well-behaved. I take this to be an attractive position; one that we will explore more.

This normalization requirement is *prima facie* equivalent to the idea that the elimination rule has to be inverse to the introduction rule. This will be covered in Chapter 5 and we'll see what sense of 'inverse' we mean here.

4.2 Dummett

Michael Dummett's analysis of logical laws and inference rules matches up nicely with this approach. There are some differences to note although I take my program to be quite similar to his. The largest difference is the context in which we take our meaning theories to be in. Where Dummett's *verificationism* is supposed to hold in general, for language, my inferentialism is restricted to mathematical and

logical contexts. This creates a sort of tension with respect to the rules considered. However, enough is said in Dummett about the logical case that I will elaborate on exactly what I take from his work.

In [34], Dummett discusses the plausibility and justification of logical laws and we'll include the inference rules there for convenience. Dummett makes an important distinction in his analysis between compositional and holistic theories of meaning. The meaning of expressions (in whatever units you please) in a holistic theory depends on the meaning of all the other expressions in a given set. We can be global holists with respect to meaning where the meaning of a sentence depends on the entirety of the language. Although this seems an incredible claim, Dummett points out that its bite comes from a denial that a meaning theory is possible. At least it denies the sort of meaning theory that compositional theories would exemplify. Compositional theories, as the name suggests, assert that the meaning of a sentence, for example, is a product of the meaning of its constituent parts. This allows for analyzing wholes in terms of parts instead of the other way around as with the holist.

But where there is opposition of this kind, there are sure to be intermediates. These intermediate theories would fall into something like a *molecular* theory of meaning.[5] That is, there are clusters of language that act as the wholes with respect to which the meaning of the parts are specified. For purposes of this paper, I take it that mathematics is mostly an autonomous language game. There is always danger is using a word like "autonomous" when it is clear that in many respects mathematics is not autonomous. The sense of autonomy that I want to use here only says that the meaning of mathematical statements only depends at most on all other mathematical statements. The cognitive and historical fact that we enter mathematical language from non-mathematical language does not contradict this. But the meaning of mathematical concepts does not *necessarily* depend on the meaning of non-mathematical concepts. So perhaps we can be molecularists in our meaning theory and take mathematics as a molecule.

A major theme of this work is a resistance to externalist programs of justification. I take this externalist requirement to be the underlying problem in the Ladyman and Presnell justificatory framework. It is also exemplified in many accounts of mathematical foundations, especially with respect to categorical foundations as in [81] and [40].[6] I resist deriving our criterion of justification from a privileged vocabulary of intuitive and *pre-mathematical* notions. It is rather widely assumed that an analysis of the foundations of mathematics requires establishing some supportive structure that is *extra*-mathematical. The circularity

[5] My position is largely in line with that expressed in [142] with respect to the meaning theory.

[6] Indeed when revisiting Feferman's [40] in a special issue of the *The Review of Symbolic Logic* [41], the foundational concerns of the participants were often externalist, as I'd characterize them.

worry looms large in considerations of foundations and it is understandable to wish for external justification. But the common mentality of "we have to start somewhere" with respect to the axioms of logic, thought, or perception often lead to assuming there is a foundations to be found or established. And yet the desire to always have justification at hand for the axioms of one system in terms of another system contradicts this dictum. A molecularist or holistic approach to linguistic meaning allows the justification story to be without foundation. It may require more in the way of explanations on how we learn the language and how it was constructed in the first place; it does *not* require a special story with respect to the axioms or logical laws.[7] As Dummett points out,

> A holistic conception of language therefore does not merely *allow* an arbitrary stipulation of the logical laws to be regarded as holding: such a conception is *demanded* by the claim that they may be arbitrarily stipulated. [34, p. 230]

Given our discussion in the previous section, we might wonder how we are to reconcile this claim with our goal of letting the introduction rules be arbitrary but the elimination rules constrained. If holism is the only option with arbitrary logical laws, but we want something like semi-arbitrary laws, can we safely land in molecularism? I think we can. To see how this is possible, we give a more detailed look at the type of meaning theory this project relies on. Dummett called the meaning theory that located the meaning of a sentence in the set of canonical warrants for its assertion *verificationism*.[8] That is, the meaning of a sentence is the 'standard', or canonical, way of verifying that sentence's truth. He called the related meaning theory that takes the set of consequences of an assertion as constitutive of its meaning *pragmatism*. This is what the listener can infer from the assertion of a sentence. The program described in the next section can be seen as the combination of Dummett's verificationism and pragmatism. And it is to this program we now turn.

4.3 An Inferentialist Treatment

In this section, I explain the inferential basis on which the next chapters will be based. We can take inferentialism as the philosophical underpinning for the

[7] Neil Tennant has explored such an explanation on how we arrived at logic through evolutionary processes in [150].

[8] I think Martin-Löf best expresses the distinction between Dummett's *verificationism* and the verificationism of the 1930s and the logical positivists. In Dummett and Martin-Löf's theory, "what is equated with verifiability is not the meaningfulness but the truth of a proposition, and what qualifies as a method of verification is a proof that the proposition is true" [92, p. 26].

category-theoretic work and conceptual analysis to follow. We focus on two aspects with respect to inferentialism: The relevance to the current project and the way it will justify my analysis of identity in Chapter 6.

First, however, we outline what inferentialism asserts as a meaning theory. Roughly, inferentialism says that the content of a concept is determined by the inferential role that concept plays in the language. This is in contrast to *representationalist* theories that require the meaning of a concept to be in reference to something (mostly) language independent. We can imagine a theory that says the meaning of a concept is its denotation or reference. The canonical example is usually taken to be physical properties. If we want to know the content of the concept "redness", we might naïvely hope that it is sufficient to group all the red things together and say that "redness" is that which all elements of this class have in common. We might further require that this thing they have in common has to be somewhat explanatory of their having this property. It wouldn't do, it seems, if all the red things happened also to be round things. It would be a mistake in this case to say that what we now call roundness is actually part of redness by an *accident* of the extension of "redness".

This harkens back to an essentialism about properties. Essentialism in this sense would say properties are those things which classes have in common. Naturally, this strong and simple sort of essentialism runs into rather large problems. But it serves as an example of the sort of external reference that representationalist theories exemplify. I do not think it would behoove us to go into too much detail about rival meaning theories here. I take it as sufficient to explain my position without too much in the way of contrast. Of course, I cannot avoid this completely.

Inferentialism plays on a larger pragmatic theory of meaning. Often cited for this theory is Wittgenstein's idea of "meaning as use" [161]. This does little in the way of defending a pragmatic position, but it gives a very broad idea of where I stand. The idea, as opposed to an externalist view, is that linguistic expressions have meanings only explicable in terms of their use in a language game. Moves in a language game provide the framework in which the meaning of expressions is fixed. What it means to fix a meaning will be dealt with below, but for now it is interesting to note a consequence of a pragmatic meaning theory. We need not require meanings to *represent* some preconceived notion of that concept. We do not need to know redness in order to check whether our concept "redness" captures it correctly. There need be no such check in a pragmatic meaning theory.

This will help us avoid what I take to be the downfall of Ladyman, Presnell, and Mayberry's argument. Specifically, it allows us to skirt the question of whether the identity type captures a *pre-mathematical* notion of identity. We may be interested in such a question for certain other purposes but the justification of the identity type does not require it. Although not explicit in their work, it seems

that the *pre-mathematical* requirement expresses a desire that the mathematical language game be representative of some non-mathematical language game. This is precisely how I understand representationalist theories. In general, they require the meaning or justification of concepts in one language game be expressed with reference to another, sometimes privileged, language game. It is here that I depart from Ladyman and Presnell by endorsing a pragmatic meaning theory instead. Thus the analysis and justification of a concept is internal to a language game and the requirement that it be reducible or 'grounded' in another seems to me to miss the point. It would at least require a preference over languages so that mathematics is necessarily dependent on basic, pre-mathematical intuitions. It is this requirement that is left unargued for; I'd suggest that we leave it so.

The nature of what this 'inferentially determined conceptual content' is and what is intended by 'inferential role' need to be further explicated. It seems question begging to define conceptual content to be that which is fixed by inferential roles. Although this is not itself the argument, it is not far off. The inferentialist asks what more could be 'in' a concept than the grounds for and the consequences of using that concept in language. If there is meaning to a concept which is not used in the warrant or a consequence of its ascription, we might wonder how we are to find this remainder. If we are talking about inferentialism for natural languages, the common point of contention is with empirical language-entry and behavioral language-exit moves. But we are not talking about the broad context here; we are talking about mathematical language.

It is an interesting question whether mathematics has similar language-entry and language-exit moves. Tennant gives three accounts that he finds possible situations for mathematics' entry and exit moves in [149, p. 9]. Since language-entry moves are perpection or observation, in the empirical case, we might see the analogue in mathematics as 'perceiving' a proof as a proof. This entry move requires a concept-wielding mathematician to use. The exit move would be to use the proof in an assertion. Tennant specifies that it would be a "speech act of *assertion backed by proof*". This option has some merit, but I find the exit move a bit odd given the entrance. Should it not be the presentation of a proof, for others to identify or not *as* a proof?

A second option is that mathematics, being closely related to the empirical sciences, has similar entry and exit moves. In sympathy with Field's [43], Tennant outlines a scientific analogue of the entry and exit moves for mathematical language. The expression of scientific laws in mathematics would be the entry and the testing of mathematically (or merely deductively) derived predictions in empirical tests would be the language-exit moves. Although I share little sympathy with Field's project, these are serviceable analogues for the language of mathematics. It might not be appropriate for all of mathematics, but mathematical science and

scientific mathematics might benefit from such an analysis.

The last option is that there are no language-entry or language-exit moves. This would be amenable, Tennant says, to Prawitz and Dummett in that it allows mathematics and logic to be extended to language fragments that require language-entrance and exit moves. I don't think this is much needed for the Prawitz or Dummett line of argument, but it still remains my favorite option. At the very least I plan to remain agnostic as of now to the fleshing out of the inferentialist moves that are not language-language moves.

4.4 Brandom's Inferentialism

To the best of my knowledge, Robert Brandom coined the term "inferentialism" [109]. But there's an irony in using something like Brandom's inferentialism for a type theory. After all, Brandom wants to capture what it is for discursive animals—sapient, not merely sentient—to use language in a meaningful way. Arguments like Searle's in [134] are classical in thinking that computers surely can never get this kind of semantics nor even remotely close to Brandom's 'sapience'. Without this capacity, we wonder if type theories might be inappropriate for such a natural linguistic theory. This difference does, however, allow us to avoid the difficulties for inferentialism in natural language and empirical statements. Many concerns regarding inferentialism are thus avoided, since we deal with the type theory instead. But then we have to justify using inferentialism at all. If we're not using it for natural languages, isn't it misleading to use the word at all? I'd say no. Inferentialism is for all of language between sapient/discursive animals. And at this moment, all evidence points to the fact mathematicians too are sapient in this way.[9] Mathematics, I want to say, is a special case of inferentialism in general. Whereas Brandom wants to hold *strong-inferentialism*, we might be able to hold that in this 'special case' or 'narrow scope', we can be *hyper-inferentialists*. It shouldn't be surprising that different parts of language, like mathematics, display different degrees of inferentialism. But when it is stronger, it remains in line with the general program of inferentialism. Much how theorems that cover large swaths of territory, in mathematics, have cases where their statements could be significantly strengthened (logically), but to preserve generality, we sometimes state the weaker one. I see my use of inferentialism as analogous to this way of arguing for weaker conclusions in general, yet allowing stronger conclusions to apply to

[9] Although in jest we might consider whether mathematicians are really sapient, there is the question about whether they are sapient *qua* mathematicians. This is not so obvious, but I think it is sufficiently clear that mathematics talk is as norm-regulated, as creative, and as reason-based as natural language. These seem enough to imply that mathematicians *qua* mathematicians are sapient in the relevant ways.

subsets of the general scope.

I will give here a short account of the three sorts of inferentialisms. Brandom, in several places[10], delineates three sorts: *weak*, *strong*, and *hyper* -inferentialism. For completeness, I include weak-inferentialism even though I do not consider it a contender currently. Weak-inferentialism states only that inferential articulation is a "*necessary* aspect of conceptual content" [19, p. 219, note 4]. Both strong and hyper inferentialism accept this; they add when it is also sufficient. Strong-inferentialism allows inferential aspects, *broadly* construed, to be sufficient for conceptual content. "Broadly" here means that sense perception-language moves and language-action[11] moves are included as inferential. Hyper-inferentialism claims that inferential aspects, *narrowly* construed, to be sufficient for conceptual content. This, of course, means that language-language moves are sufficient by themselves, to account for conceptual content.

I hold that indeed, the language-language moves are by themselves sufficient. But only in the context of mathematics. Even Brandom admits that hyper-inferentialism is plausible "at most for some abstract mathematical concepts" [19, p. 220, note 4]. This is exactly what I am pursuing here. Of course, if we want to interact with the world in concrete ways, we'll need to allow both perception-language and language-action moves as 'broadly inferential'.[12]

The pragmatism here emphasizes that knowledge of language is more a know-how than know-that or even that there is no difference between the two as Martin-Löf prefers. This is naturally aligned with mechanization. After all, if we take seriously the claims about a computer's ability to have semantics of the representationalist kind [134], then we should be glad that language is process-based, not internal state-based. Know-how is what a computer can do. Know-how is much the same as input-output behavior. Type theory, as any formal logical calculus, is a way of keeping score in the game of giving and asking for reasons, in the narrow realm of mathematical discourse or maybe just type-theoretic discourse. The rules are the score cards, and the computers are the umpires [73]. Computers make sure everything happens just right, like chess programs.

Coming now to the concept at hand, the meaning of identity, under a broadly inferentialist framework, is characterized by the moves to and from the concept and ascription of that concept in a language game. Two questions need to be answered before a conception of identity is settled: Claiming two things to be equal, what are we to infer from this ascription? And under what conditions are we entitled to claim that two things are equal?

These two questions will be answered with respect to MLTT. I will draw, in

[10]Here I take the discussion in [19] to be the easiest to understand and thus I use its characterizations.

[11]These terms come from Sellars in [140].

[12]See Chapter 4 of [21].

part, from the recent work of Brandom in expanding and interpreting some of Sellar's work. Some parallels will be drawn between type theory and Brandom's discussion of cross-sortal identity. This is intended to help tie the current discussion of HoTT to current philosophical work in philosophy of language, logic, and metaphysics. This is not an attempt to legitimize HoTT by its relation to philosophy, but to place the questions we ask of HoTT in a different context. This move is primarily against work that takes HoTT in a vacuum and treats it as if much work has not been done concerning similar concepts.

In Section 2.2, we pointed out that terms can only be of one type. This can be viewed as saying that there are no cross-type identities. For our purposes, we'll use the notion of cross-sortal identities instead. This notion of 'cross-sortal identity' comes in two flavors. If two sortals (synonyms include 'kinds', 'intensional classes', and '*types*') are such that one has a weaker criterion of identity than the other, we might have *weak* cross-sortal identity. The example is that a kitten can be identical to a cat. Although different intensional classes, they have the appropriately related criteria of identity. In fact being the same kitten implies being the same cat. The converse is true if we modify it to stipulate that the same cat, if it is a kitten, is the same kitten. This is what is meant by a 'weaker' criterion of identity. The criterion of application of 'kitten' and 'cat' are different, but their identity criteria are the 'same' [20, p. 218]. Brandom also gives an example where the criteria of application is the 'same' but the criterion of identity is not. But some sortals do not have such related criteria of identity.

In intensional type theory, like MLTT, each term can only be a member of one type. It makes no sense to make a judgment like $(a, a) : A \times B$ unless we have $A \equiv B$ as a definitional equality. The interpretation of Brandom's phrase "strong cross-sortal identity" is thus an underpinning notion in intensional type theory. Without careful work, no given term is allowed to be a member of definitionally non-identical types. Brandom holds that there are no strongly cross-sortal identities. Of course, if we admit objects such as sub-types, then we clearly allow cross-sortal identities, but we call these *weakly* cross-sortal identities.

A short qualification is in order. The following arguments interpret Brandom's work, and some of Sellars, but not on their level. They are not discussing formal languages and it pays off to be careful about where you are talking about formal and natural languages. So any exceptions or criticisms I give to their arguments are to be viewed as a product of the change of venue. Whether their arguments, where applicable, will be successful in this formal framework remains to be seen. So when I say that there are exceptions to Brandom's conclusion that there are no strongly cross-sortal identities, this is not a challenge to the conclusion since the contexts we work in are relevantly different. It will be fruitful to view the following as an exploration and expansion of the interpretive work in [20] in contexts not

considered by the author.

Another centrap aspect of Brandom's program is that of *assertion* or *proposition*. The difference between proposition and judgment has been explored above in Section 2.2.1 only with respect to equalities. But of course, these are fundamental categories in the discussion of the foundational status of MLTT and HoTT. With my characterization of inferentialism, I use the word "ascription" when describing the way that concepts get used in inference. Another way is to call it "assertion". Importantly, Brandom distinguishes between assertions and stipulations. In discussions, some have suggested that propositions (types) are assertions and judgments are stipulations. This counters the very motivation for creating Martin-Löf Type Theory. The suggested parallel would entail that giving a term of a type is proving an assertion and yet demonstrating that it is indeed a proof would be a 'mere' stipulation. Stipulations do not have *force* for Brandom, and so do not entail a commitment to the proposition that is expressed by it. I do not think this properly captures the notion we want.

Martin-Löf spends a fair bit of time analyzing the nature of judgment and proposition. Martin-Löf initially defines propositions as those entities that the logical operators operate on and assertions as those that we prove and "appear as premises and conclusions of logical inference" [92, p. 2]. It is actually the judgments then that have the assertory force for Martin-Löf, and not the propositions. In many presentations of logic, the proposition A implicitly asserts the truth of A. A key aspect of MLTT is that judgments — in a semi-Kantian sense of the word — are those above and below an inference line. Logical judgments, for Martin-Löf are of the form

$$A \text{ is a proposition, } A \text{ is true, } A \text{ is not true.}$$

And for our type theory, we have an important judgment "A `Type`". In modern logic texts — Martin-Löf points out that this is through Frege's influence — we use the assertion (judgment) "A is true". But for this to make sense, we have to know that A is of an appropriate character — that is, that A is a proposition. Thus Martin-Löf points out a need to make explicit the condition that "A is a proposition" is known or judged prior to considering "A is true".

To use a concrete example, this implicit premise in inference rules of the propositional nature of the symbols is made explicit for $\vee$-introduction.

$$\frac{A\ prop \qquad B\ prop \qquad A\ true}{A \vee B\ true}$$

We represent in the above the *assertion* "A is a proposition" with $A\ prop$. Now, it is important that Martin-Löf's considerations of propositions and judgments are colored by philosophical intuitionism. In [92], Martin-Löf makes the explicit claim

that to judge, in the sense above, is the same as to know. This is where the epistemic force comes in. For comparison, Brandom attempts to recover classical logic in his theory of inferentialism, what Tennant calls a 'quietist' perspective [149]. But without the connection between knowledge and assertion that intuitionism affords, many have questioned the ability of Brandom's theory to bear the distinction between assertion and stipulation. Martin-Löf has a more direct way of saying why assertions have epistemic force: The equivalence between knowledge and judgments with proof (or evidence). Therefore the parallel between MLTT and Brandom's theory becomes the following. Assertions, those with *force*, are proved judgements in MLTT. Stipulations, those without *force*, are unproven or assumed judgments. Martin-Löf requires that contexts (antecedents of sequents) are judgments but they need not be evident or proven; contexts are therefore stipulations for Brandom.

We notice that the notion of proposition is internal to the logic while the notion of judgment is the framework in which we get assertion, stipulation, and proof. So propositions in MLTT are those entities that the rules of inference govern and are inductively defined with logical operators. Judgments contain and are about propositions and either assert or stipulate them. A lovely presentation and elaboration of Martin-Löf's work along these lines is given by one of his students, Johan Georg Granström, in [53].

4.5 Meaning and Truth

The concepts of meaning and truth have both found themselves at the center of contemporary analytic philosophy. Long before even the 'linguistic turn' [16] in philosophy, a preoccupation with meaning and truth has been a certain norm. No philosophy can wholly avoid the question and neither do I here. I will say that although my remarks will be relatively summary, they are supposed to support the work on identity that can be found in the next chapters.

I take as influential the works of Michael Dummett, especially in *The Logical Basis of Metaphysics*. Although my program is not whole-cloth taken from Dummett's work, it would be beneficial to indicate the relevant parts for the current project. The relevant parts are perhaps unsurprisingly with respect to meaning. The verificationist and inferentialist programs are, in a direct sense, contrasted with Tarski semantics. Where Tarski's schema fixes meaning to evaluate truth predicates, a Dummettian or inferentialist analysis will fix truth (or truth preservation) to explain meaning.

It is a relevant bit of sociology that professional mathematicians generally equivocate truth with (having a) proof, especially in non-formal systems. We can make sense of this by affirming that meaning depends on truth, not the other

way around. Both proof-theoretic semantics and the inferential program used here are ways in which to express this relationship. When we take inferential moves — which implicitly require a notion of validity and truth — as the basis for our meaning theory, we find that meaning can be fixed in a mathematically appropriate way. Mathematical logic, and proof theory, give us the tools we need to talk about 'inferential roles', 'inferential moves', and of course 'proof' in a formal system. In Chapter 6, I use the categorical representation of mathematical logic to pin down more general notions of inference.

In order to explicate how meaning can be determined (or 'fixed') via inferential roles and moves, I need to first explain the success criterion for my theory. One of the first criterion that a meaning theory traditionally needs to satisfy is some sort of reason why the analysis is that of *meaning* at all. This is a significant challenge for representationalist and similar programs. There should be a way to indicate that whatever philosophy of meaning one holds is indeed about *meaning*. Under representationalist or platonic philosophies, a challenge much like the naturalistic fallacy can be levied. They may give us elaborate ways to think of meaning and show painstakingly that it coincides with our intuitive ascription of meaning to expressions. But we are left with the question of whether what the representationalist (or other non-pragmatic philosophies) posits *is* meaning or simply correlated with it.

This challenge is only well-formed when we imagine the meaning to be in a sense independent of the use of the expression. Explaining that the theory matches our use is only a part of the program for representationalists; it is to give credibility to the theory, not necessarily justificatory merit. If we opt for a pragmatic view of meaning, we immediately avoid such a challenge. For it does not make sense to follow up the statement of a pragmatic meaning theory with a question of whether it really captures *meaning*. For the pragmatic theory, there is no meaning that is not dependent on use. There is no independent meaning that our analysis has to 'capture', and so the criterion is lifted.

Of course we must then ask why the pragmatic theory can hold that meaning is so dependent and neatly avoid this criticism. But this too presupposes an anti-pragmatic view of language. As if meaning is independent of use in some respect and the pragmatist has to glue this remainder back into use. I take it as an unconvincing challenge that pragmatic meaning theories fail to account for meaning 'proper'. At the theoretic level, then, a pragmatic theory of meaning seems question begging. If the justification relies on a denial of the independence of meaning and use while at the same time an affirmation of the dependence of meaning and use, we seem to have a confusion. But the denial of this sort of independence is a meta-theoretic claim.

The sorts of questions we allow in our metatheorizing exclude those of the

foundationalist and the representationalist. So while it is true that meaning-as-use as a thesis can seem question begging to the representationlist, I find no reason to avoid this consequence. To put it even more polemically, I'd side with certain 'therapeutic' views like those of Wittgenstein and Rorty [125] when it comes to the appropriate questions asked of a meaning-theory.

A criterion that I think *does* need to be satisfied is an explication of identity and non-identity of meaning between expressions. The inferential programs do so. Two expressions are identical when they have the same inferential role. Two expressions are different when their inferential roles are different. For the identity of meaning, then, we pass the buck to the identity of inferential roles.

I believe this definition of the identity of meaning would not work in the full natural language setting — or at least I think it is a substantially different project.[13] As our context is helpfully constrained to mathematics and type theory, it is perhaps more reasonable that synonymy be better defined. Likewise, the motivation for considering inferential roles to constitute the criterion of identity among (the meaning of) expressions stems from the fact that mathematics has a amenable 'space of reasons'. The question of meaning then becomes how we are to fix positions in such a space of reasons.

The space of reasons[14] for Sellars, does rely on a logic of reasons in the language game and so we need a logic in which mathematics is discursive. It happens that mathematics has a ready logic of reasons: proof. It is tricky to pin down 'reasons' in natural language because we get subtleties of explanation and justification, but for mathematics if someone asks why X asserts Y, we can simply point to X having a proof of Y. In this way, we get relatively more grip on where concepts are located in the space of reason.

[13] Of course there are prominent challenges to theories that claim to fix synonymy. Something like the analytic/synthetic distinction might well creep in. I do take the challenges of Quine [120] seriously and add that by changing the discussion to logical synonymy mathematics narrowly, we avoid the charges expressed there.

[14] Sellars coins this lovely phrase in *Empiricism and the Philosophy of Mind* with "The essential point is that in characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says" [136, § 36].

5 | Harmony

The concept of Harmony is not new but I will give it some new clothing. It seems that the idea has cropped in many distinct places somewhat independently. I will not be able to give an elaborate history of the notion, but I will go over some modern conceptions that I think deserve a unification. My goal is to unify the various conceptions of Harmony. Indeed, the notion I will put forward will easily and uniformly show the Harmony of the standard rules for the following connectives:

$$\&, \vee, \rightarrow, \forall, \exists, \neg, \bot, \text{and } \top$$

The conception of Harmony I will create draws from categorical logic and is sufficient to establish the other sorts of Harmony; importantly, it will not only allow me to establish the Harmony of the above connectives but also the identity rules of MLTT. My Harmony will be given by adjunctions between functors of appropriate types. We will be giving this new characterization below.

For most purposes, then, my conception will unify the different 'harmony's in the literature. I take this unification to have several advantages. Immediately, we have the satisfaction of the unification of several apparently different concepts. Logicians and mathematicians tend to prefer having one concept that is 'powerful' and widely applicable. For us, this is adjointness and will enable us to bring several concepts under one. But more importantly, adjoints do not simply include harmonious properties, they are central for other parts of category theory. It is a surprising fact that this simple relationship between two functors is so fundamental to the logic and structure of mathematics. It strikes us as impressive that adjoints are fundamental in the category theory even before I show that they give us Harmony as well.

To give an analogy to why this is important, I take an example from mathematical physics. The notion of group was found to be fundamental to many algebraic structures that were commonly employed in mathematics: usually number systems with a binary operation and inverse. But it was impressive (and Wigner would say "unreasonably" so [157]) that group theory turned out to be so good at representing certain aspects quantum field theory. The fundamental interactions of particle

physics were captured by the mathematical tools from algebra even when the algebras were not designed to do so, at least not explicitly. After all, the definition of a group was useful in mathematics but there was little if any indication that it would be a powerful empirical tool. Similarly, the definition of adjunct was useful in category theory and several algebraic fields. But its usefulness in describing logical notions was not, as it were, baked in. Largely due to Lawvere did we see the connection between the logic we knew and loved and adjunctions in category theory. This is just one more connection between category theory and logic (properly even the *philosophy* of logic) in a long list of them. The deeper and more general ties will be given in Chapter 6 where we put logic in category-theoretic contexts.

Another advantage of unifying these notions under the concept of adjoints relates to inferentialism. The punchline of this chapter is that

Concepts determined by adjunctions are harmonious.

You'll notice the slight ambiguity there since we've only encountered Harmony, albeit very briefly, with respect to *rules*, not concepts. But this is precisely the point. Since conjunction is determined by adjunction, for example, its rules are harmonious. This changes the question of Harmony to focus on the concepts that are fixed by adjunctions, not on free-floating concepts and rules. When I say 'fixed', I mean to indicate that the adjunctions give us *unique* characterizations of concepts like conjunction. This is along the inferentialist lines given in Chapter 4 in that we look at concepts by their rules and at rules by their ability to fix concepts uniquely.

For now, we focus on the novel connection: *Harmony as adjoints*. The connection is clearest in type theory since the rules are a bit more elaborate — consider the inclusion of β and η rules[1]as well as the fact that *objects* and not simply propositions are inferred. However, in this section I will generalize the connection to natural deduction and related logics. This will put us in a position to establish the Harmony of the identity rules in MLTT. And this will be the internal justification of Path Induction; the rules of identity are harmonious.

5.1 Conceptions of Harmony

I want to get clear here on what Harmony is, at least how I use it. To do this, I will sketch the various conceptions of Harmony that are in the literature, distinguishing and relating my Harmony to these other conceptions along the way. The history

[1]These will be discussed below.

is important as it follows rather lock-step with the development of key logical innovations.

Gerhard Gentzen is often attributed with coming up with a general concept of Harmony especially in how Prawitz cashes this out — Prawitz explicitly cites Gentzen as an inspiration [115, p. 75]. I include here the translation that explains the position that Gentzen, and thus Prawitz, exposits:

> By making these ideas more precise it should be possible to display the E-inferences as unique functions of their corresponding I-inferences, on the basis of certain requirements. [146, p. 81]

This is clearly reminiscent of Section 4.1's conclusion with Prawitz and Dummett. I do see Gentzen expressing an asymmetric approach to the rules of inference in this quote. The introduction rules are meaning creations on their connective that need not represent any pre-theoretical notion. The elimination rules must however be tailor-made, as it were, to the introduction rules. Prawitz calls this sort of balance between the elimination and the introduction rules the *inversion principle*:

> Observe that the elimination rule is, in a sense, the inverse of the corresponding introduction rule: by an application of an elimination rule one essentially only restores what had already been established if the major premiss of the application was inferred by an application of an introduction rule. This relationship between the introduction rules and the elimination rules is roughly expressed by the following principle, which I shall call the *inversion principle*. [114, p. 33]

I reproduce the inversion principle itself to show how Prawitz believes we check this inversion.

> Let α be an application of an elimination rule that has B as a consequence. Then, deductions that satisfy the sufficient condition [...] for deriving the major premiss of α, when combined with deductions of the minor premisses of α (if any), already 'contain' a deduction of B; the deduction of B is thus obtainable directly from the given deductions without the addition of α. [114, p. 33]

In other words, Prawitz's Harmony can be established by showing that detour formulae are locally reducible. This is part of the argument for the conception of Harmony used in this paper. I take the point to be the sense in which the elimination rules are inverses of the corresponding introduction rules with a corresponding test. This inversion principle, although modified for a different context below, will

reflect the basic idea that Harmony is about inverse relationships between complementary sets of inference rules.

Here we come to a distinction between conceptions of Harmony. Some authors, including Dummett [34] and Steinberger [144] consider harmony to entail only that the elimination rule is inverse to the introduction rule. These authors usually call the converse claim, that the introduction rule is inverse to the elimination rule, "stability". I will use "Harmony" to be the conjunction of these two claims.

For different characterizations of Harmony, we return in part to the history of justifying rules of inference. First we recall the position of Nuel Belnap when responding to Prior's tonk example. Belnap required that a connective's rules ensure a conservative extension of the fragment of the logic that doesn't include that connective or rules. Further, in the style of inferentialism explained in the last chapter, we also need to verify that the meaning of a connective is properly fixed by its rules.[2] This is a uniqueness condition on the connective. To say what uniqueness means, let σ be a connective with rules of inference $\mathcal{R}$. Then we say that $\mathcal{R}$ ensures the uniqueness of σ' when for any new connective σ', we can show that when we replace σ with σ' in all rules of $\mathcal{R}$, the following is derivable $\sigma(x_1, \ldots, x_n) \leftrightarrow \sigma'(x_1, \ldots, x_n)$. In other words, if σ and σ' have the same rules, modulo symbol choice, then the connectives are logically equivalent. Of course, if this did not hold, then we have denied inferentialism in logic, for two concepts would have the same rules governing them and yet be somehow distinguishable.

We maintain this uniqueness condition in a structural form for our conception in 5.2. It will be the universality of adjunctions that implies the appropriate notion of uniqueness of a connective, namely identity. The fixing of meaning will not be with respect to a biconditional but with respect to isomorphisms. Although in the classical case this will turn out to be the same, the treatment with adjoints allows us to also make sense of proof-relevant uniqueness.

The point to notice for us is that in order to use the inferential framework, we must check that the rules of inference governing a connective create a (relevantly) unique inferential role of that connective. The inferential role is determined by seeing whether something can have the same rules but mean something different. This test will be easiest in categorical setting, to which we leave it for now.

We return to Belnap's main requirement: that of conservative extension. Dummett phrases harmony in a general sense as follows:

> harmony [...] obtains between the verification-conditions or applications-conditions of a given expression and the consequences of applying it cannot, by appealing to its conventionally accepted application-conditions and then invoking the conventional consequences of apply-

[2] In [32], the authors explicitly say that when rules are harmonious they are 'meaning-conferring'. This, essentially, is the point of Harmony.

> ing it, establish as true some statement which we should have no other means of establishing: in other words, when the language is, in a transferred sense, a *conservative extension* of what remains of it when the given expression is deleted from its vocabulary. [34, p. 247, my emphasis]

We note that having rules that provide a conservative extension is not intrinsic to the connective those rules govern. It depends on the context and logic fragment that we're considering. It is trivial but perhaps illustrative that anything is a conservative extension of an inconsistent logic. I take this to be the sort of curiosity that obfuscates the conservative extension from providing a motivated requirement for Harmony. The fact that this conservative extension requirement is so context-dependent makes me hesitant to accept it as the definition of Harmony. But it would not surprise me that my requirement is equivalent to this conservationism.

Closer to our concerns in MLTT, we have a computer science take on Harmony. In their discussion of constructive modal logics, Pfenning and Davies in [110] use different terms. We will borrow these terms when discussing the identity rules of MLTT. To represent that the elimination rules are inverse to the introduction rules, we use the term *local soundness*. To represent that the introduction rules are inverse to the elimination rules, we use the term *local completeness*.

We define these concepts more precisely as follows:

Local Soundness The elimination rules cannot be too strong. No matter how we apply elimination rules to the result of an introduction rule we cannot gain any new information. We demonstrate this by showing that we can find a more direct proof of the conclusion of the elimination which does not first introduce and then eliminate the connective in question. This is witnessed by a *local reduction* of the given introduction and the subsequent elimination. This is exactly what we showed in the case of $\wedge$ above.

Local Completeness The elimination rules are not too weak. There has to always be a way to apply elimination rules so that we can reconstitute a proof of the original proposition from the results by applying introduction rules. This is witnessed by a *local expansion* of an arbitrary given derivation into some eliminations followed by some introductions. This property for conjunction follows.[3]

I reproduce the local reduction for conjunction that was presented in Section 4.1 that, in the new terminology, shows the local soundness of conjunction:

[3] Another way to express this is by focusing on the strength of the elimination rules. For instance, Steinberger says that disharmony can happen two ways: "E-strong disharmony" and "E-weak disharmony" [143, p. 1].

$$\dfrac{\dfrac{\begin{array}{cc}\mathcal{D}_1 & \mathcal{D}_2 \\ A & B\end{array}}{A \wedge B}\,\wedge\mathrm{I}}{A}\,\wedge\mathrm{E1} \quad \text{reduces to} \quad \begin{array}{c}\mathcal{D}_1 \\ A\end{array}$$

We now also show the local completeness of conjunction with a local expansion.

$$\begin{array}{c}\mathcal{D} \\ A \wedge B\end{array} \quad \text{expands to} \quad \dfrac{\dfrac{\mathcal{D}}{A \wedge B}\,\wedge\mathrm{E1} \quad \dfrac{\mathcal{D}}{A \wedge B}\,\wedge\mathrm{E2}}{A \wedge B}\,\wedge\mathrm{I}$$

Local soundness and completeness together constitute *Harmony* for conjunction in the Gentzen-style natural deduction. Another way to say this is that we showed harmony and stability for conjunction. A slightly different way to express Harmony in the case of type theory is the following.

Local Soundness The elimination rule is the inverse of the introduction rule. If we eliminate on a term that was introduced via a type's introduction rule, we should get the same thing back.

Local Completeness The introduction rule is inverse to the elimination rule. This is a uniqueness condition for terms of a type.

Fortunately for the type theorist, Harmony in this sense is made precise by the notion of η (local completeness) and β (local soundness) conversion. We will use the case of binary products that serve as a reinterpretation on the discussion of conjunction. Here are the slightly abbreviated rules for binary products:

First the formation rule:

$$\times \text{ Formation } \frac{A\ \texttt{Type} \qquad B\ \texttt{Type}}{A \times B\ \texttt{Type}}$$

And as expected, the canonical terms of type $A \times B$ are pairs

$$\times \text{ Introduction } \frac{a : A \qquad b : B}{\langle a, b\rangle : A \times B}$$

We also get two maps such that

$$\times \text{ Elimination } \quad \frac{p : A \times B}{\pi_0(p) : A} \qquad \frac{p : A \times B}{\pi_1(p) : B}$$

where π_0 and π_1 are the first and second projections respectively. Now we have the β rules

$$\times\beta\ \frac{\langle a, b\rangle : A \times B}{\pi_0(\langle a, b\rangle) \equiv a : A} \qquad \frac{\langle a, b\rangle : A \times B}{\pi_1(\langle a, b\rangle) \equiv b : B}$$

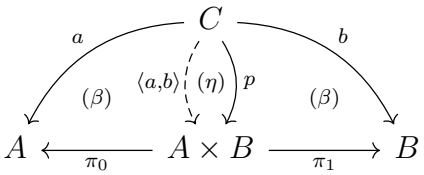
And here we have the η rule

$$\times\eta \quad \frac{p : A \times B}{p \equiv \langle \pi_0(p), \pi_1(p) \rangle : A \times B}$$

The β and η rules are crucial here. The β rule says that if you project out of a canonical term of $A \times B$, namely a pair, you get the appropriate argument back. Importantly, the elimination rule only tells you *where* the projection lives, the β rule makes the stronger claim that we know exactly *what* the projection is, if the argument is a term that we got from $\times$ Introduction. Here we see that the β rule stipulates that the eliminations, i.e. the projections, are an inverse to the introduction rule, the pairing terms.

The η rule is a uniqueness condition. It is the converse to the β rule in the sense that it says that the introduction rule is inverse to the elimination rule.

As we will later state harmony and identity in category-theoretic terms, it may be helpful to express the β and η rules in a categorical setting. The main point to take away from this exposition is that the universal nature of (binary) products ensures Harmony. Keeping with our conjunction/product example, we present the Harmony of binary products. First, the diagram.



In words: given any $a : C \to A$ and $b : C \to B$, there exists a unique arrow $\langle a, b \rangle : C \to A \times B$ such that the left and right triangles commute. I have marked the inside of the triangles with (β) because the commutativity of the triangle is expressed by the β rule for products. After all, if we state this equationally, we get

$$\pi_0 \circ \langle a, b \rangle = a$$

which is to say $\pi_0(\langle a, b \rangle) \equiv a$ in our type theory. Similarly $\pi_1(\langle a, b \rangle) \equiv b$. These are precisely the β rules.

The η rule for products expresses their uniqueness. More explicitly, for any $p : C \to A \times B$ that also makes the triangles commute, that is $\pi_0 \circ p = a$ and $\pi_1 \circ p = b$, we have $p = \langle a, b \rangle$. And so in our language, this says that $p \equiv \langle \pi_0(p), \pi_1(p) \rangle$, which is precisely the η rule for products.

The example of binary products illustrates what the β and η rules accomplish. They are the formalization of Harmony, local soundness and local completeness respectively. Not all the rules have the same kind of Harmony, but they are created with Harmony already included. Where the rules of natural deduction include only

an introduction and elimination part, the rules of Martin-Löf type theory include also β and usually η rules explicitly.

5.2 Adjoints

In order for our slogan "Harmony as adjoints" to be clear, I need to give a short description of what *adjoints* are. Adjoints are a crucial concept in both category theory and categorical logic. The uses and power of this concept are hard to overstate. Indeed, to get a feeling for how important they are for logical purposes, we note that the following are determined by adjoint situations: $\wedge, \vee, \top, \bot, \supset, \forall x\colon A, \exists x : A, =$, comprehension, quotients, weakening, and contraction. All of these are important for a logical type theory. For our purposes, we will focus on equality as an adjoint, but first I want to point out why adjoints are important for our particular project.

Adjoints express Harmony. Suppose a concept F can be expressed by a functor that is left adjoint to a functor G. Then the following one-to-one correspondence — or *two way inference rule* — will hold:

$$\frac{F(X) \to Y}{\overline{\overline{X \to G(Y)}}}$$

This can be read to say that for every arrow from $F(X)$ to Y there exists a unique arrow from X to $G(Y)$ and vice versa.

For example, in the category of all small categories, we have a product functor $\times : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ that is right adjoint to the diagonal functor $\delta : \mathbb{C} \to \mathbb{C} \times \mathbb{C}$. We can express the relationship between these two functors with a two way inference rule.

$$\frac{\delta(A) \to \langle B, C\rangle}{\overline{\overline{A \to \times(\langle B, C\rangle)}}} \qquad \text{or more familiarly} \qquad \frac{\langle A, A\rangle \to \langle B, C\rangle}{\overline{\overline{A \to B \times C}}}$$

The forward (downward) direction gives us the introduction rule for products. Reading arrows into an object as terms of that type, we can read this off as follows. Given two arrows $(b : A \to B, c : A \to C)$, we have an arrow we call $\langle b, c\rangle : A \to B \times C$. Another way to say the same thing is that given $b\colon B$ and $c\colon C$ we have a term we call $\langle b, c\rangle\colon B \times C$. This is the introduction rule for binary products.

The backward (upward) direction gives us the elimination rule for binary products. Similar to the introduction rule, we can read this direction off as follows. Given an arrow $p : A \to B \times C$, we have two arrows $\pi_0 : B \times C \to B$ and $\pi_1 : B \times C \to C$. Equivalently, this gives us the rule that given $p\colon B \times C$, we have

two terms $\pi_0(p):B$ and $\pi_1(p):C$. The reason we have p as a parameter is to keep the interpretation of terms of types over a specific type A. Thus in the category theory, we simply compose the p with a projection to get an arrow from A to B and A to C. This is the elimination rule for binary products.

Importantly, there is a morphism θ that witnesses the bidirectionality of the adjunction. That is,

$$\theta : \mathrm{Hom}_{\mathbb{C}\times\mathbb{C}}(\delta(-),-) \to \mathrm{Hom}_{\mathbb{C}}(-,\times(-))$$

is a natural isomorphism. This means that for objects $A, B, C \in \mathbb{C}$, we know that $\theta_{A,(B,C)}$ is an instance of the introduction rule for binary products explained above. Similarly, since θ is an isomorphism for any arguments, we have θ^{-1} as well. And $\theta^{-1}_{A,(B,C)}$ is an instance of the elimination rule for binary products. Importantly we have the following facts for any objects $A, B, C \in \mathbb{C}$

$$\theta_{A,(B,C)} \circ \theta^{-1}_{A,(B,C)} = 1 \quad \text{and} \quad \theta^{-1}_{A,(B,C)} \circ \theta_{A,(B,C)} = 1.$$

This permits us to say that the logical counterpart to products, conjunctions, is harmonious, or has harmonious rules. After all, $\theta_{A,(B,C)}$ is one instance of the introduction rule schema for products (conjunctions). Similarly, $\theta^{-1}_{A,(B,C)}$ is a pair that constitutes an instance of the elimination rule schema. Thus, their being inverses shows that the introduction followed by an elimination is the identity and the elimination followed by an introduction is also the identity. And now we can see clearly how adjointness expresses Harmony, at least for the case of products. Adjunctions are natural isomorphisms and so as long as both directions of the adjunction correspond to the introduction and elimination rules, we have Harmony.

Since the rules governing a connective are inverses if the connective is determined by adjunction, showing that identity is an adjunction will give us the internal justification of the identity rules. We will give such a characterization next chapter but already we see that it will be sufficient to give a well-known adjunction for identity in order to ensure the Harmony of its rules. One benefit of this conception of Harmony is that the 'information' in the definition of local soundness and completeness is replaced by an explicit inverse relationship between the rules. But we can also establish that adjunctions ensure that the elimination rules are not too strong or too weak, which is Tennant's preferred way of expressing his harmony [149].

6 | Lawvere's Law

In this section we go over the necessary category theory to establish what we'll be calling "Lawvere's Law". We give this name to the conception of identity that follows in order to connect it with Leibniz's Laws of identity: the identity of indiscernibles and the indiscernibility of identicals. We might formalize this law thus:

$$\forall P\,(Px \leftrightarrow Py) \leftrightarrow x = y \qquad \text{(Leibniz's Law)}$$

As Russell hints at in [127, ch. 4,5], one of the problems with this law is that it focuses on properties. The attractive alternative is to focus on relations; after all, identity is a *relation*. Mathematically speaking, Leibniz's Law does not afford us much general use. After all, Leibniz and his contemporary readers did not use the identity of indiscernibles to talk about abstract objects. It was only with respect to substance, as Leibniz defines that concept, that the Law was *contingently* true. It is odd that in many discussions of identity in mathematics, we still start at this point in the conversation.

I will thus leave this Law to be considered in other contexts. In mathematics, we do not consider this definition of equality as the only legitimate option, although it is not a useless one. In set theory, for instance, we defined equality on a set to be interpreted with the model-theoretic object $\{\langle x, x\rangle \mid x \in I\}$ where I is the set underlying the interpretation. But since we're in a type theory we cannot use this standard interpretation of equality. It is instructive, however, to consider why this interpretation is common. This is the *smallest* reflexive set whose pairs include all those elements of the underlying set I. This treatment suggests that identity is the smallest reflexive relation on $I \times I$. This is much closer to the sort of conception of identity that will be given below, but we will avoid this classical model-theoretic sort of reasoning.

The method of using category theory to argue for identity's proper formulation takes justification. I will shortly defend my choice to use categorical logic in this paper with two lines of argument. The first is that a very important part of the HoTT project is giving categorical semantics for the type theory. So right off the bat, we have the close relationship between category theory — and more specifi-

cally categorical logic[1]— and HoTT. The second point to make is that categorical logic is the generalization of many of the classical (i.e. traditional and Tarskian, not necessarily as opposed to intuitionistic) cases. Quantifiers as adjoints, and many other revelations in categorical logic suggest that to get a universal characterization of a mathematical concept, like identity, it does well to look to the category theory.

I do not broach the foundational topic of whether category theory is a good or autonomous foundation for mathematics as in [81]. I do, however, think that most parties agree that the *language* of category theory, at least, does provide conceptual tools that are most helpful in clarifying mathematical concepts. This is all I will say with respect to why category theory makes such a prominent appearance in the following.

The way the identity type is described in Section 2.2.2 gives the technical detail but does not describe the intuition behind *why* the identity type is constructed this way. In Martin-Löf's prominent [91], the identity type is given a 'reflection principle' between definitional and propositional identity. The introduction of terms in the identity type is given by definitional or intensional equality:

$$\frac{a \equiv_A b}{r : \mathtt{Id}_{\mathtt{A}}(a, b)}$$

Martin-Löf gives a meaning explanation of this identity type:

> We now have to explain how to form canonical elements of $\mathtt{Id}_{\mathtt{A}}(a, b)$. The standard way to know that $\mathtt{Id}_{\mathtt{A}}(a, b)$ is true is to have $a \equiv_A b$. Thus the introduction rule is simply: if $a \equiv_A b$, then there is a canonical proof r of $\mathtt{Id}_{\mathtt{A}}(a, b)$. Here r does not depend on a, b, or A; it does not matter what canonical element $\mathtt{Id}_{\mathtt{A}}(a, b)$ has when $a \equiv_A b$, as long as it has one. [91, p. 60, changed notation]

But this identity reflection gives us uniqueness of identity proofs — the UIP discussed briefly near the end of Section 3.1. The 'I-elimination' is given in our notation as

$$\frac{c : \mathtt{Id}_{\mathtt{A}}(a, b)}{a \equiv_A b}$$

which is surely stronger than the one we use since we distinguish between propositional and definitional identity in our type theory. But we don't need such a strong rule, though it has uses in other domains. For instance, in Martin-Löf's

[1] Maybe cite work on cubical sets and lawvere duality to bipointed sets as a way of modeling HoTT. At least discuss the alternatives and hopes of the community.

work on inductive definitions [90], he defines identity as inductively generated by the reflexivity rule, which is how our identity type is understood.

This reminds us of the uniqueness principle of identity types that Ladyman and Presnell use in their work. That is, terms of the identity type are not definitionally equal to `refl` (UIP) but they are *propositionally* equal (uniqueness principle). Licata has given a similar exposition of the identity type and makes the emphasis that "the real force of J is to assert that PathFrom M is inductively defined, with generator (M , Refl)", which is to reiterate that the identity type is inductively generated by `refl`, as desired.[2] This is the sort of meaning explanation that I think is appropriate for the identity rules. The rules as given in this paper bear out this sort of interpretation and so we have our explanation of the identity rules. But for the justification, we must answer the question "Why should we expect that identity is inductively generated by reflexivity?" This is the question to which we now turn. We will give an independent analysis of identity that is general and includes the classical one and show that it is indeed also generated from reflexivity.

6.1 Hyperdoctrines

In this section we cover the notion of *hyperdoctrines*, introduced famously in [77]. More importantly for our current project is the work Lawvere did in [79]. The precise formulation of hyperdoctrines has changed since these papers have been published, but only in presentation. We will use the more modern presentations in [65], [111], and [112].

First we say what a first-order signature will consist of.

Definition 6.1. A *first-order signature* $\Sigma = \langle \mathcal{S}, \mathcal{F}, \mathcal{R} \rangle$ consists of

- A set $\mathcal{S}$ of *types* (or sorts) $A, B, \ldots$;
- A set $\mathcal{F}$ of *function symbols* $f, g, \ldots$ together with a type for each symbol such as $f : A_1 \times \cdots \times A_n \to B$;
- A set $\mathcal{R}$ of *relation symbols* $R, S, \ldots$ together with a type for each symbol such as $R : A_1 \times \cdots \times A_n$,

Definition 6.2. The *terms* over a signature Σ are defined inductively as follows

- If x is a variable of a type A, then $x : A$ is a term;

[2] Dan Licata's illuminating post at http://homotopytypetheory.org/2011/04/10/just-kidding-understanding-identity-elimination-in-homotopy-type-theory/ shows how he arrives as this conclusion and what he means by PathFrom etc.

- If $f : A_1 \times \cdots \times A_n \to B$ is a function symbol and $t_1 : A_1, \ldots, t_n : A_n$ are terms in their appropriate types, then $f(t_1, \ldots, t_n)$ is a term of type B.

But in general we have terms in contexts of a list of variables. We denote a term t in context $\vec{x}$ by

$$\vec{x} \mid t$$

where $\vec{x}$ is a finite list of distinct variables that contains all the variables in t.

We will also need to specify what the *formulae-in-context* are. We relativize formulae to the signature. The set $\mathbf{Form}(\Sigma)$ of first-order formulae over a signature Σ is built up inductively as follows:

- Atomic formulae: $R(t_1, \ldots, t_n)$ where $R \in \mathcal{R}$ is a relation symbol with type $B_1 \times \cdots \times B_n$ and each $t_i : B_i$ for $1 \leq i \leq n$.
- $\mathbf{Form}(\Sigma)$ is closed under $\vee, \wedge, \to$, and $\neg$ starting with the Atomic formulae.
- We also have $\top$ as a formula.

Since these formulae can have free variables, we denote a formula ϕ in context of a (typed) variable list $\vec{x}$ that includes all the variables free in ϕ by

$$\vec{x} \mid \phi$$

To interpret the terms over a signature in categorical logic, we interpret this signature in a category.

Definition 6.3. A *categorical interpretation* for a collection of terms-in-context consists of

- A category $\mathbb{C}$ with finite products.
- A mapping $[\![-]\!] : \mathcal{S} \to \mathbb{C}$ assigning an object $[\![A]\!]$ to each type A.
- A mapping $[\![-]\!] : \mathcal{F} \to \mathbb{C}$ assigning an arrow $[\![f]\!]$ to each function symbol F such that if $f : A_1 \times \cdots \times A_n \to B$ then $[\![f]\!] : [\![A_1]\!] \times \cdots \times [\![A_n]\!] \to [\![B]\!]$ in $\mathbb{C}$.

We note that constants in a signature are function symbols with no domain type. Thus a constant c might be $\to B$. This is interpreted categorically by $[\![c]\!] : 1 \to [\![B]\!]$ where 1 is the terminal object of $\mathbb{C}$.

We now define the general notion of hyperdoctrine.

Definition 6.4. A *hyperdoctrine* is a categorical structure that acts as the target for the interpretation of a first-order language over a signature. It has the following:

- A base category $\mathbb{C}$ with finite products.

- A (pseudo)functor $\mathcal{P} : \mathbb{C}^{\text{op}} \to \mathcal{CCC}$ where $\mathcal{CCC}$ is the category of Cartesian Closed Categories.[3] This means that for any object $A \in \mathbb{C}$, there is a CCC $\mathcal{P}(A)$. And for each arrow $f : B \to A$ in $\mathbb{C}$, a functor between Cartesian closed categories $\mathcal{P}(f) : \mathcal{P}(A) \to \mathcal{P}(B)$. We will abbreviate $\mathcal{P}(f)$ as f^*. We will also call f^* the substitution morphism. (A pseudofunctor is a functor that preserves composition and identities only up to a coherent choice of isomorphism. [12, p. 47])

- For all $f : B \to A$, we require the functor $f^* : \mathcal{P}(A) \to \mathcal{P}(B)$ to have a right and left adjoint, denoted by $\forall_f$ and $\exists_f$ respectively.

  For the adjoints $\forall_f$ and $\exists_f$ the following *Beck-Chevalley* conditions hold: for all $f : A \to B$ and $g : B \to C \in \mathbb{C}$,

  $$f^* \circ \exists_f \cong \exists_f \circ (1_A \times f)^* \qquad \text{(Beck-Chevalley)}$$

  likewise for $\forall_f$.

As a demonstration of the power of adjoints in this context, we can characterize the Cartesian structure of $\mathbb{C}$ simply by adjoints. The terminal object (the functor $1 : \mathbf{1} \to \mathbb{C}$) is given as a right adjoint to the unique morphism $! : \mathbb{C} \to \mathbf{1}$. We can express this as a two-way rule of inference as before, with $\star$ the unique element of the terminal category 1:

$$\frac{!(A) \to \star}{A \to 1(\star)}$$

or more simply

$$\frac{\star \to \star}{A \to 1_{\mathbb{C}}}$$

Since the morphism on the top is the unique morphism of the terminal category, it is always satisfied, which gives us the fact that there is always a unique morphism $A \to 1_{\mathbb{C}}$ for every $A \in \mathbb{C}$. Notice that for every object A of $\mathbb{C}$, the category $\mathcal{P}(A)$ is also Cartesian closed. We think of the terminal object of $\mathcal{P}(A)$ for some A to be the always true predicate.

Finite products are derived similarly. The product functor is right adjoint to the diagonal functor $\delta : \mathbb{C} \to \mathbb{C} \times \mathbb{C}$ that can be intuitively thought of as a doubling up of arguments $X \mapsto \langle X, X \rangle$. So we get the adjunction

[3] We could also have this category be $b\mathcal{CCC}$, the bi-Cartesian Closed Categories. This would allow us to have an initial object (false) and sums in general. But for our purposes we will assume the less restrictive $\mathcal{CCC}$.

$$\frac{\delta(A) \to \langle B, C \rangle}{A \to \times(B, C)}$$

which we can rewrite for clarity as

$$\frac{\langle A, A \rangle \to \langle B, C \rangle}{A \to B \times C}$$

And we saw in Section 5.2 how this adjunction gives us the usual properties of products (and logical conjunctions). For any $\mathcal{P}(A)$, we consider the products to represent conjunction.

Finally, the exponents of a Cartesian Closed Category are also given by adjunction. The exponent functor $(-)^A$, for any $A \in \mathbb{C}$, is right adjoint to the product functor $A \times (-)$. This gives us the following adjunction scheme

$$\frac{A \times B \to X}{B \to X^A}$$

The exponentials of any $\mathcal{P}(A)$ corresponds with implication which we'll represent with $\Rightarrow$. Perhaps importantly, if we take B to be B^A itself, we get the adjunction

$$\frac{A \times B^A \to B}{B^A \to B^A}$$

We call the upper morphism the evaluation morphism because given two arguments $\langle a : A, f : A \to B \rangle$, it evaluates f at a, returning an element of B. If we're working with the category $\mathcal{P}(A)$ for some A, the arguments to the evaluation morphism are $\langle Q, Q \to R \rangle$ and results in R with $Q, R \in \mathcal{P}(A)$. This is precisely modus ponens! We can also rewrite the arguments in a proof-relevant way. Given $a : A$ and $f : A \to B$, we have a proof $f(a) : B$.

But the interpretation of hyperdoctrines, as above presented is still to be defined.

Definition 6.5. A *categorical interpretation* for the first-order language over Σ consists of:

- A hyperdoctrine $(\mathbb{C}, \mathcal{P})$;
- An interpretation of terms-in-context over Σ into $\mathbb{C}$;
- A mapping $[\![-]\!] : \mathcal{R} \to \mathcal{P}$ that assigns to each relation symbol $R \in \mathcal{R}$ an element of the Cartesian Closed Category $\mathcal{P}([\![B_1]\!] \times \cdots \times [\![B_m]\!])$ where the type of R is $B_1 \times \cdots \times B_m$.

From this, we can build up an interpretation of formulae-in-context $\vec{x} \mid \phi$:

- $[\![\vec{x} \mid s = t]\!] = \langle [\![\vec{x} \mid s]\!], [\![\vec{x} \mid t]\!] \rangle^*(\delta_{[\![A]\!]})$ where $s : A$ and $t : A$;
- $[\![\vec{x} \mid R(t_1, \ldots, t_n)]\!] = \langle [\![\vec{x} \mid t_1]\!], \ldots, [\![\vec{x} \mid t_n]\!] \rangle^*([\![R]\!])$;
- $[\![\exists y\, (\vec{x} \mid \phi)]\!] = \exists_{[\![A]\!],I}([\![\vec{x}, y \mid \phi]\!])$ where $\vec{x} : A_1 \times \cdots \times A_n$ and $y : A$. Here, $I = \prod_{i=1}^{n} [\![A_i]\!]$ to make the subscripts more legible;
- The case of $[\![\forall y\, (\vec{x} \mid \phi)]\!]$ is analogous to the existential case.
- The Cartesian structure, namely the products, of each $\mathcal{P}(A)$ gives us the usual interpretation of $\wedge$.

The codomain of the pseudo-functor $\mathcal{P}$ varies among authors. We can take it to be a partially ordered set, which is geared toward classical (proof-irrelevant) interpretations. This would make $\mathcal{P}$ functorial as well. Andrew Pitts takes this as an illustrative *case* of hyperdoctrine, though he does not define hyperdoctrines with this requirement, in [111, p. 41]. We follow Lawvere's original specification of hyperdoctrines, where he only requires cartesian closed categories for the codomain of $\mathcal{P}$. This requirement is too strong if we take logic fragments instead, but we will not explore this detail much here.

It is helpful to consider the terminology that Lawvere used in describing hyperdoctrines. While they are only names, they do hint as the ideas they are intended to capture. The collections of 'types' and 'terms' suggests the modeling of type theory in a hyperdoctrine. The Cartesian Closed Category $\mathcal{P}(X)$ for any object X of $\mathbb{C}$ is called the "attributes of type X" which we might even read as the collection of generalized predicates over X. The arrows in $\mathcal{P}(X)$ are then called "deductions over X". This name hints at why we don't restrict ourselves to a pre-order codomain since this would imply the uniqueness of deductions, making our theory proof-irrelevant. Importantly, given an arrow $f : B \to A$ in $\mathbb{C}$, if we apply f^* to some $X \in \mathcal{P}(A)$, we call $f^*(X)$ the "substitution of f in X".

It should be noted here that the morphisms that we specify $\exists$ and $\forall$ as adjoints to can also vary among authors. Some authors, especially if they take the codomain of $\mathcal{P}$ to be a pre-order, will only define the quantifiers as adjoints to some projections. But since we assume that our deductions have left and right adjoints, this requirement is not needed. So we define the quantifiers in a more general way than if we had a pre-order. The usual quantifiers are special cases of our more general ones. But it is of interest that the identity predicate will be an instance of the general existential.

The intuition behind a hyperdoctrine can be derived from the following two examples:

Example 1. Consider the types to be arbitrary sets, terms as arbitrary functions between sets, and $\mathcal{P}$ to be the powerset functor $\mathcal{P}(-)$.

Within this example, we can see more directly what f^* would do for a given $f \in \mathbb{C}$. If we recognize that the arrows between $\mathcal{P}(X)$ (the deductions over X) are really just subset relations, then we have the following mapping $f^* : \mathcal{P}(A) \to \mathcal{P}(B)$ for some $f : B \to A$:

$$X \in \mathcal{P}(A) \overset{f^*}{\longmapsto} \{b \mid f(b) \in X\} \in \mathcal{P}(B)$$

or in other words, f^* is the inverse image operator on powersets. We also notice that since $\mathcal{P}(X)$ for any X is a partial order, there is at most one arrow between any two sets. This forces the deductions (the arrows the codomain of $\mathcal{P}$) to be proof-*ir*relevant.

Two special cases of this 'substitution' map f^* are weakening and contraction. Weakening is a substitution along a projection. That is, for $\pi_1 : A \times B \to A$,

$$\begin{aligned} \pi_1^* : \mathcal{P}(A) &\to \mathcal{P}(A \times B) \\ X &\longmapsto \{(x, b) \mid x \in X \text{ and } b \in B\} \end{aligned}$$

This can be seen as adding a variable from B to a predicate on A. Contraction is substitution along a diagonal $\delta : A \to A \times A$. As expected, its action is to collapse two variables into one:

$$\begin{aligned} \delta^* : \mathcal{P}(A \times A) &\to \mathcal{P}(A) \\ X &\longmapsto \{a \in A \mid \langle a, a \rangle \in X\} \end{aligned}$$

The usual quantifiers, as specified above are adjoints to weakening π^* as indicated in the important expression

$$\exists \dashv \pi^* \dashv \forall$$

This states that the existential quantifier is left adjoint to weakening and the universal quantifier is right adjoint to weakening.

Equality in this example will be the classical relation $Eq(A) = \{\langle a, a \rangle \in A \times A\}$ which is certainly not the identity type we are trying to capture. Although this example helps us grasp the structure of a hyperdoctrine, we do not need to take this to be the canonical or standard example. A proof-relevant hyperdoctrine, as the next example, will provide us with a clearer understanding of how the identity type correlates to the categorical conception of identity.

Example 2. This next example will allow a full Hom-set of deductions between attributes; this makes the logic proof-relevant. Thus this is an example that comes closer

to the spirit of HoTT.

We take the same types and terms as in the first example, namely arbitrary sets and mappings: the category $\mathbb{C} = \mathbf{Sets}$. But we define $\mathcal{P}$ differently. For any $A \in \mathbb{C}$, we have that $\mathcal{P}(A)$ is the category of "sets over A". Or, as we'll say, the slice category $\mathbf{Sets}/A$. The objects of this category are arrows $f : X \to A$ in $\mathbf{Sets}$. The arrows in $\mathbf{Sets}/A$ are arrows $g : X \to X'$ in $\mathbf{Sets}$ such that the following triangle commutes:

$$\begin{array}{ccc} X & \xrightarrow{g} & X' \\ & \searrow^{f} & \downarrow{\scriptstyle f'} \\ & & A \end{array}$$

Given $g : B \to A$ in $\mathbf{Sets}$, we define $g^* : \mathbf{Sets}/A = \mathcal{P}(A) \to \mathbf{Sets}/B = \mathcal{P}(B)$ as follows: Given the following situation with $\varphi \in \mathbf{Sets}/A$,

$$\begin{array}{ccc} & & Y \\ & & \downarrow{\scriptstyle \varphi} \\ B & \xrightarrow[g]{} & A \end{array}$$

we want an object in $\mathbf{Sets}/B$ which is the morphism given by the pullback of g along φ.

$$\begin{array}{ccc} B \times_A Y & \longrightarrow & Y \\ {\scriptstyle b}\downarrow & \lrcorner & \downarrow{\scriptstyle \varphi} \\ B & \xrightarrow[g]{} & A \end{array}$$

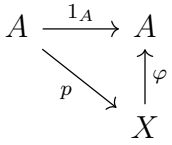
where b is the object of $\mathcal{P}(B)$ that we desired. This is where pseudo-functorality is important, as pullbacks are only unique up to isomorphism.

Thus, for $f : A \to B$, we have the left adjoint to f^* given by composition. That is, $\exists_f : \mathcal{P}(A) \to \mathcal{P}(B)$ is given by composing $g \in \mathcal{P}(A)$ with f:

$$X \xrightarrow{g} A \xrightarrow{f} B$$

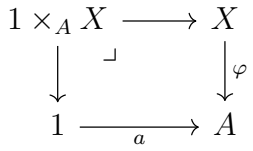
We might call deductions of the form $1_A \to \varphi$ in $\mathcal{P}(A)$ "proofs" of φ. But this lets us associate proofs of propositions with points (terms) of the types. Proofs in this sense are *sections* of formulas in $\mathcal{P}(A)$, then. Diagrammatically, a proof p of a formula φ is represented for this hyperdoctrine as the

commutative diagram

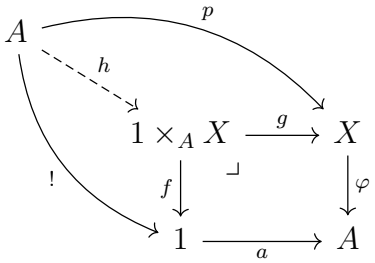


or equationally $\varphi \circ p = 1_A$.

With this formulation of 'proof', given $a : 1 \to A$ and $\varphi : X \to A$, we have the diagram



And using the universal properties of pullbacks and terminal objects, we extend this to



which gives us a proof, h, of $a \circ f \in \mathcal{P}(A)$. We know that h is a proof of $a \circ f$ by 'diagram chasing' and the fact that p is a proof of φ:

$$\begin{aligned}(a \circ f) \circ h &= (\varphi \circ g) \circ h \\ &= \varphi \circ (g \circ h) \\ &= \varphi \circ p \\ &= 1_A\end{aligned}$$

showing that h is a section of $a \circ f$. So we can assign a proof h of $a \circ f$ for each a and where f is the projection from the pullback. In this way, a proof of $\exists_f(\varphi)$ is a mapping that assigns to every y a single x such that $f(x) = y$ but also includes a proof of $\varphi^*(x)$. This matches up closely with Σ types in type theory. We could also generalize proof here to include 'premises' other than the always true one 1_A.

The notion of hyperdoctrine is given largely in terms of adjointness, thus this conception of logic is more general than the classical notion of model might suggest.

For Lawvere, the fact that cartesian closure, equality, and the quantifiers can all be expressed and determined by adjointness speaks the importance of the concept. Indeed, the fact that the classical quantifiers are adjoints to simple context weakening is an elegant fact often pointed to in demonstrating the centrality of the concept of adjoint. For our purposes, however, we will focus below on the equality predicate introduced as an adjoint to the (morphism assigned to the) diagonal morphism.

6.2 Equality in a hyperdoctrine

The punchline for this section is a characterization of identity that goes like this

> Identity is that relation that implies all and only reflexive relations.

But to arrive at this requires a categorical characterization in the framework of hyperdoctrines as described above.

Equality predicates are, not surprisingly, defined as adjoints to simple functors. Our equality predicate on a type $A \in \mathbb{C}$ which we write as Eq_A will be left adjoint to the substitution along the diagonal. In general, the diagonal $\delta_A : A \to A \times A$ gives rise to the the substitution $\delta_A^* : \mathcal{P}(A \times A) \to \mathcal{P}(A)$ and will be called the contraction functor. We define the equality attribute (relation or predicate) on $A \times A$ to be left adjoint to δ_A^* at the terminal object of $\mathcal{P}(A)$ which we call $\top$. As in our definition of the hyperdoctrine, we express this with a generalized existential quantifier.

For a substitution morphism like $\delta_A^* : \mathcal{P}(A \times A) \to \mathcal{P}(A)$, we have a left adjoint $\exists_{\delta_A} : \mathcal{P}(A) \to \mathcal{P}(A \times A)$. Since $\mathcal{P}(A)$ is cartesian closed, it has a terminal object $\top$ which can be thought of as the 'always true predicate'. We define the equality relation to be $\exists_{\delta_A}$ applied to this always true predicate. That is,

$$Eq_A := \exists_{\delta_A}(\top) \tag{6.1}$$

For clarity, the following picture shows us how Eq_A is derived given the diagonal morphism:

$$\begin{array}{c} A \\ \downarrow{\scriptstyle \delta_A} \\ A \times A \end{array} \qquad \text{gives us} \qquad \begin{array}{ccc} & & \mathcal{P}(A) \\ & \exists_{\delta_A} \swarrow & \dashv \quad \uparrow{\scriptstyle \delta_A^*} \\ & & \mathcal{P}(A \times A) \end{array}$$

And notice that $Eq_A \in \mathcal{P}(A \times A)$ as defined above. Putting this in our familiar inference rule format, we get the following biconditional for any relation $R \in \mathcal{P}(A \times A)$:

$$\frac{\exists_{\delta_A}(\top) \to R}{\top \to \delta_A^* R} \text{ or } \frac{Eq_A \to R}{\top \to \delta_A^*(R)} \quad \begin{array}{c} \text{in } \mathcal{P}(A \times A) \\ \text{in } \mathcal{P}(A) \end{array}$$

This can be read as "Identity on A implies the relation R if and only if ..." and so we see how this gives an inferentially specified sufficient and necessary condition for the identity relation. What then does the bottom line say? Since we read arrows as inferences or deductions in our categories of attributes $\mathcal{P}(A)$, the lower inference says the always true predicate $\top$ implies the contraction of R. The contraction functor δ_A^* replaces two variables by one. Since R is a relation on A, contracting it takes $R(x, y)$ and gives $R(x, x)$ (or of course $R(y, y)$). But since $\top \to R(x, x)$ in this case, we have that R is reflexive under trivially true assumptions.

Thus we can rewrite the adjunction to get our characterization of identity:

$$\frac{Eq_A(x, y) \to R(x, y)}{\top \to R(x, x)}$$

In our usual notation we have the easily readable

$$\frac{Eq_A(x, y) \vdash R(x, y)}{\vdash R(x, x)}$$

The forward (downward) entailment can be read as saying that "if equality implies R, then R is reflexive". The backward (upward) entailment can be read as saying that "if R is reflexive, then equality implies R". Both of these put together give us our unique characterization of identity:

Theorem 6.1 (Lawvere's Law)**.** *Identity is that relation that implies all and only reflexive relations.*

It is unique because adjoints are unique up to isomorphism and an isomorphism here would be logical equivalence.

This is the central point of this chapter and indeed much of this paper. This characterization of identity has all the marks that I have been arguing are necessary for a mathematically and philosophically sound meaning statement. It's inferentially stated, since it simply calls identity that inferential role that has a certain characteristic, namely implying all and only reflexive relations. As just mentioned, the fact that it's an adjunction gives us the relevant sort of uniqueness needed to 'fix' meaning in the inferential sense. And finally, adjunctions also give us harmony. So if the rules of identity are determined by this characterization, then they are harmonious.

To emphasize, this is the *mathematical* notion of identity, generally characterized. It is neither pre-mathematical nor specific to the classical model-theoretic

notion of identity. The strength of this characterization needs further explication and I will prove the consequences of characterizing equality by this "Lawvere's Law".

Theorem 6.2. *Equality is reflexive.*

Proof. If we take R to be Eq_A, then we get

$$\frac{Eq_A(x,y) \vdash Eq_A(x,y)}{\vdash Eq_A(x,x)}$$

whose top line is always satisfied by the identity morphism in $\mathcal{P}(A \times A)$. □

Theorem 6.3. *Equality is symmetric.*

Proof. The fact that contraction can contract either variable gives us a rather simple proof. This can be shown by the simple derivation

$$\frac{\dfrac{Eq_A(x,y) \vdash R(x,y)}{\vdash R(y,y)}}{Eq_A(x,y) \vdash R(y,x)}$$

And letting R be Eq_A shows us that $Eq_A(x,y) \vdash Eq_A(y,x)$. □

Theorem 6.4. *Equality is also transitive.*

Proof. Since $\mathcal{P}(A)$ is cartesian closed, we have exponentials corresponding to implications. Assume that $Eq_A(a,b)$ and $Eq_A(b,c)$. Choosing R to be $Eq_A(y,c) \Rightarrow Eq_A(x,c)$, we get

$$\frac{Eq_A(x,y) \vdash Eq_A(x,c) \Rightarrow Eq_A(y,c)}{\vdash Eq_A(x,c) \Rightarrow Eq_A(x,c)}$$

But the bottom is trivially satisfied by the identity morphism of $Eq_A(x,c)$ in $\mathcal{P}(A \times A)$. So the adjunction ensures that $Eq_A(x,y) \vdash Eq_A(x,c) \Rightarrow Eq_A(y,c)$ and replacing the variables to get $Eq_A(a,b) \vdash Eq_A(a,c) \Rightarrow Eq_A(b,c)$ which means that,

$$Eq_A(a,b), Eq_A(b,c) \vdash Eq_A(a,c)$$

as desired. □

Finally, we prove that Lawvere's Law gives us the principle of substitution *salva veritate*. This will require a proof closer to the hyperdoctrine notation above. We will use the two projections which are defined with respect to the product adjunction. That is, we have $\pi_1 : X \times Y \to X$ and $\pi_2 : X \times Y \to Y$ in a finite product category $\mathbb{C}$. These give rise to $\pi_1^* : \mathcal{P}(X) \to \mathcal{P}(X \times Y)$ and $\pi_2^* : \mathcal{P}(Y) \to$

$\mathcal{P}(X \times Y)$ which are the weakening functors, adding variables. As described in the examples of the last section, weakening and contraction are inverses and so $\delta_X^* \circ \pi_1^* = \delta_X^* \circ \pi_2^* = 1_{\mathcal{P}(X)}$.

We also need the fact that substitution functors behave well with composition. Given $f : X \to Y$ and $g : Y \to Z$ in a category of types $\mathbb{C}$ and an attribute $\phi \in \mathcal{P}(Z)$, we assume

$$f^*(g^*(\phi)) \cong (g \circ f)^*(\phi)$$

Theorem 6.5. *In a hyperdoctrine $(\mathbb{C}, \mathcal{P})$, for $f : A \to B$ and predicates $\phi, \psi \in \mathcal{P}(B)$, if the deduction*

$$f^*(\phi \Rightarrow \psi) \to f^*(\phi) \Rightarrow f^*(\psi)$$

in $\mathcal{P}(A)$ is an isomorphism (equivalence), then for any predicate $\alpha \in \mathcal{P}(A)$, we have the deduction

$$Eq_A \to (\pi_1^*(\alpha) \Rightarrow \pi_2^*(\alpha))$$

in $\mathcal{P}(A \times A)$.

Proof. Let $\alpha \in \mathcal{P}(A)$. We have the identity deduction $\alpha \Rightarrow \alpha$ and so

$$\begin{aligned} \top &\to (\alpha \Rightarrow \alpha) \\ &\to (\pi_1 \circ \delta_A)^*(\alpha) \Rightarrow (\pi_2 \circ \delta_A)^*(\alpha) && (\delta_A^* \circ \pi_i^* = 1_{\mathcal{P}(A)}) \\ &\xrightarrow{\sim} \delta_A^*(\pi_1^*(\alpha) \Rightarrow \pi_2^*(\alpha)) && \text{(assumption)} \end{aligned}$$

And given the fact that equality is defined as the left adjoint to contraction at $\top$, we get an arrow

$$Eq_A \to (\pi_1^*(\alpha) \Rightarrow \pi_2^*(\alpha))$$

□

This deduction is the general form of the substitutivity of equality. That is, if α is a unary relation on A, then for $x, y : A$

$$Eq_A(x, y) \vdash (\alpha(x) \Rightarrow \alpha(y)).$$

This says that equality is a relation that all predicates respect. Thus, we have now proved that Lawvere's Law entails all the standard features of equality: reflexivity, symmetry, transitivity, and substitutivity. The proofs for these facts bear striking resemblance to the proofs given in Section 2.2.2 for the identity rules of MLTT. So not only do both MLTT identity and Lawvere's Law prove the above facts about identity, the way to prove them is sufficiently similar to already suggest their equivalence.

The two rules for identity are as follows:

$$\frac{\vdash R(x,x)}{Eq_A(x,y) \vdash R(x,y)}\,\mathrm{LL}$$

$$\frac{x{:}A, y{:}A, p{:}\mathtt{Id_A}(x,y) \vdash R(x,y,p)\ \mathtt{Type} \qquad x{:}A \vdash t(x){:}R(x,x,\mathtt{refl_x})}{x{:}A, y{:}A, p{:}\mathtt{Id_A}(x,y) \vdash J(t;x,y,p){:}R(x,y,p)}$$

Ignoring the `Type` declaration (the left premise) in Path Induction and making some variable contexts implicit, we can rewrite this as

$$\frac{\vdash t(x){:}R(x,x,\mathtt{refl_x})}{p{:}\mathtt{Id_A}(x,y) \vdash J(t;x,y,p){:}R(x,y,p)}$$

which is clearly similar to one direction of Lawvere's Law. I argue that their similarity is strong enough to justify the identity rules of MLTT.

The identity rules of MLTT indicate that identity is generated from reflexivity and this, I claim, is what Lawvere's Law states as well. Not only do we have the necessary features of identity (i.e. reflexivity, symmetry, transitivity, and substitutivity), they are also both the smallest reflexive relations of their respective frameworks. The identity rules of MLTT say this through Path Induction, which shows that for any (provably) reflexive relation R, a proof of identity is transformed into a proof of R. Another way to say this is that identity implies all reflexive relations in MLTT. And this is precisely what Lawvere's Law states as well.

We have established that the independent conception of identity gained by our analysis of hyperdoctrines requires the identity relation to imply all and only reflexive relations. I have called this Lawvere's Law and we saw that it implies that identity is reflexive, symmetric and transitive. We also note that this was given by an adjunction, namely to contraction at the always true predicate $\top$. Given the discussion in the last chapter, this shows us that the characterization of identity is harmonious. Moreover, the identity of MLTT was also shown to have the expected features of reflexivity, symmetry, and transitivity. Most importantly, they both interpret identity as that relation that implies all and only reflexive relations. Along with the provability of the identity of indiscernibles, this shows that the identity concepts of MLTT and hyperdoctrines rise or fall together. That is, if Lawvere's Law is justified, then so is Path Induction.

The only difference between (one direction of) Lawvere's Law and Path Induction is that Path Induction allows the relation R to depend on an identification. This means that R is not a relation on $A \times A$ but on $A \times A \times \mathtt{Id_A}$. But this is to only require a canonical and propositional notion of reflexivity. Whereas reflexivity in

a hyperdoctrine is being a relation whose contraction is deduced, in MLTT the reflexivity could have been proven in some particular way. Proving $R(x, x, \mathtt{refl}_x)$ as a premise means nothing more than that R is reflexive in the canonical way. This is a distinction that the general hyperdoctrine cannot handle. However, if we pick a proof-relevant hyperdoctrine, our equality *is* with respect to a *proof* of reflexivity, namely the trivial one.

6.3 Equivalent Identities

The general and inferentialist notion of identity given by Lawvere's Law implies the identity rules of MLTT. Clearly the introduction rule is taken care of by the proof that identity is reflexive. That identity implies all reflexive relations will give us the elimination rule of Path Induction. Moreover, the homotopical interpretation plays no part here so we may refer to it simply as Id-Elim. We'll have to introduce another adjunction to prove that Lawvere's Law implies Id-Elim: comprehension. It is perhaps fortuitous that Lawvere introduced both the equality and comprehension adjoints in the same paper, [79].

Comprehension makes a type out of a predicate. We will apply comprehension to identity in order to get the context for the motive correct. After all, the motive is not a predicate on two variables, but three, possibly depending on an identification.

We will characterize comprehension in a hyperdoctrine as an adjoint. In any hyperdoctrine, we have the functor:

$$\Sigma_{(-)}1_{(-)} : \mathbb{C}/X \to \mathcal{P}(X)$$

which is defined on objects as follows:

$$\begin{array}{c} E \\ {\scriptstyle p}\downarrow \\ X \end{array} \longmapsto \Sigma_p 1_E$$

When this functor has a right adjoint, we say the hyperdoctrine satisfies the *Comprehension Schema*. This adjoint takes predicates on X to objects over X. For a predicate ϕ in $\mathcal{P}(X)$, we denote the comprehension object by $\{X : \phi\}$. Recall that morphisms in $\mathbb{C}/X$ are commutative triangles over X. This gives us the double rule of inference

$$\begin{array}{ccc} E & & \{X:\phi\} \\ {\scriptstyle p}\downarrow & \to & \downarrow{\scriptstyle p_\phi} \\ X & & X \end{array}$$
$$\overline{\Sigma_p 1_E \to \phi}$$

We have further conditions on the comprehension adjunct. Although we will not need this, it is reasonable to require that the comprehension functor R behaves well with the Σ and geometric functors.

Given $f : X \to Y$, the following diagram commutes:

$$\begin{array}{ccc} \mathcal{P}(X) & \xrightarrow{R} & \mathbb{C}/X \\ {\scriptstyle \Sigma_f}\downarrow & & \downarrow{\scriptstyle f_!} \\ \mathcal{P}(Y) & \xrightarrow{R} & \mathbb{C}/Y \end{array}$$

which basically tells us that Σ functors commute with composition in the slice. Here, $f_!$ is the left adjoint to the substitution (change of base by pullback) f^* in the slice categories. It turns out just to be post-composition: Given $a : A \to X$, we get $f_!(a) = f \circ a$. This is as assumption that gives us structure to integrate comprehension into the existing hyperdoctrine structure.

Further we take as an axiom the following existence claim: For any $\beta : B \to A$, with $\eta : B \to R(\Sigma_\beta 1_B))$ given by the unit of comprehension adjunction, there is a point

$$1 \to \Sigma_\eta 1_B. \tag{6.2}$$

This allows us to prove the major theorem of this paper, that Path Induction is implied by equality in a hyperdoctrine with comprehension.

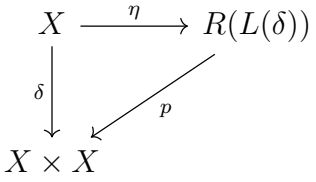
Theorem 6.6. *Lawvere's Law in a hyperdoctrine that satisfies the comprehension schema implies the validity of identity elimination in HoTT (Path Induction).*

Proof. If a hyperdoctrine satisfies the comprehension schema, as we will assume from now on, we have the following situation:

Let $X \in \mathbb{C}$, where $\mathbb{C}$ is our base category of types. To simplify notation, let $L = \Sigma_{(-)} 1_{(-)}$ and let the comprehension functor R be given by adjunction $L \dashv R$. Now take the diagonal map $\delta : X \to X \times X$ so that we have $L(\delta) = Eq_X$ in $\mathcal{P}(X \times X)$ and $R(L(\delta))$ in $\mathbb{C}/X \times X$. If we take the unit of the comprehension adjoint, we get the rule of inference

$$\frac{L(\delta) \to L(\delta)}{\delta \to R(L(\delta))}$$

which is trivially satisfied. But $\delta \to R(L(\delta))$ is a commutative triangle since it is a morphism in $\mathbb{C}/X \times X$. This gives us the diagram



Here, η is the unit of $L \dashv R$ at $L(\delta)$. Since this triangle commutes, we have $\delta = p \circ \eta$. By functorality of L, we get

$$\Sigma_\delta 1_X = \Sigma_p(\Sigma_\eta 1_X). \tag{6.3}$$

Since $L(\delta) = Eq_X$ and R is the comprehension functor, we interpret $R(L(\delta))$ as the identity type. In this way, the equality predicate is hypostatized into a type of the base category, which we will suggestively denote $\mathtt{Id}_X$.

In the structure of our hyperdoctrine with comprehension, we have assumption 6.2 that gives us, with $\delta : X \to X \times X$ and $\eta : X \to R(L(\delta))$, an arrow

$$\sigma : 1 \to L(\eta) = \Sigma_\eta 1_X$$

Context judgments like $x, y{:}A, p{:}\mathtt{Id}_\mathtt{A}(x, y) \vdash R(x, y, p)\,\mathtt{Type}$ are interpreted as objects of the slice over the object that interprets the context. So our $\eta : X \to \mathtt{Id}_X$ models the dependency on the identification when considering the premises of Id-Elim. Suppose we have a motive R in $\mathcal{P}(\mathtt{Id}_X)$ and a map (that interprets a *term*)

$$t : 1 \to \eta^*(R) \text{ in } \mathcal{P}(X) \tag{6.4}$$

Recall that $\eta^*(R)$ is the substitution of η in R. Thus, this is the interpretation of our major premise of Id-Elim, i.e. $t{:}R(x, x, \mathtt{refl}_\mathtt{x})$.

So now we have to provide a morphism for $J : 1 \to R$ in the context of identified terms $\mathcal{P}(\mathtt{ID}_X)$ for our conclusion. But since $\Sigma_\eta \dashv \eta^*$ by definition, we have the counit of the adjunction given by

$$\frac{\Sigma_\eta(\eta^*(R)) \;\to\; R}{\eta^*(R) \;\to\; \eta^*(R)}$$

This gives us an arrow $\mathrm{ep} : \Sigma_\eta(\eta^*(R)) \to R$. We also have an arrow

$$\Sigma_\eta(t) : \Sigma_\eta 1_X \to \Sigma_\eta(\eta^*(R)) \tag{6.5}$$

And since $\sigma : 1 \to \Sigma_\eta 1_X$, this establishes the arrow

$$\Sigma_\eta(t) \circ \sigma : 1 \to \Sigma_\eta 1_X \to \Sigma_\eta(\eta^*(R)) \tag{6.6}$$

Finally, composing with ep we get

$$J : 1 \xrightarrow{6.6} \Sigma_\eta(\eta^*(R)) \xrightarrow{\text{ep}} R \tag{6.7}$$

which is the term that we needed. □

7 | Conclusion

This paper has attempted to give a justification of the identity rules in Homotopy Type Theory by providing a justification without appeal to the homotopic interpretation [8]. Even though one of HoTT's main innovations is exactly this interpretation, the justification of its rules does not essentially rely on it. Opposing the contraints that Ladyman and Presnell place on the justification of the rules, I do not appeal to non-mathematical notions in order to replace the 'problematic' homotopic interpretation. Instead, I use a general and independent mathematical characterization of identity which shows that the identity rules of HoTT are founded in the properly *mathematical* notion of identity. But to do this, by removing the homotopic interpretation, I refer to the identity rules of the formal type theory of Martin-Löf in order to show this match between a mathematical identity and the identity rules, especially identity elimination. These more basic identity rules, those of Martin-Löf Type Theory, are justified by a two step method: first by justifying their internal coherence and then justifying their ability to fix and determine the meaning of identity. The order of these steps is, of course, not important. I have attempted to sketch reasons for accepting this method of justification by referring to a meaning theory for mathematics that is inferentialist.

The fact that I use a mathematical justification of the identity rules seems to conflict with the avoidance of notions of homotopy since this aversion seems motivated by the mathematical nature of homotopy theory. So why do I avoid homotopy, like Ladyman and Presnell desire, and yet still give a mathematical justification of the identity type's rules? The answer is twofold. The most prominent reason to avoid the use of homotopy is to give an *independent* notion of identity and show that the identity rules capture that notion. This independence is a weak sort that is not trying to be non-mathematical, as has been repeatedly stressed. The way I see it, we are giving a meaning explanation — as described in the introduction — which confirms that the identity rules do what we want them to. To decide what we want the rules to do, we draw from general mathematical experience to guide our notion of identity. In this way, we hold fixed most of mathematics, and verify that the addition of the identity rules of MLTT appropriately respects our notions. This can be viewed as an admissibility claim for the rules of

inference, but not to some logical framework, unless the logicists are right. Instead it's showing that the identity rules are conceptually consistent with mathematics; that it does not disrupt the way mathematics 'hangs together'.

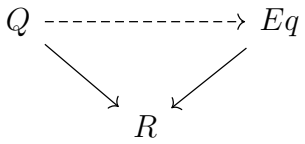
Besides this sort of admissibility claim, we have another reason for leaving homotopic justification out of the preceding pages while at the same time remaining mathematical. I feel that those who are seriously concerned with the homotopical interpretation of a foundations would have written off any attempt at using homotopy itself to justify the use of homotopy. Closely related to the motivation of independence described above, I believe that the general logical notions in hyperdoctrines provide a *neutral* ground in which to examine the structure of identity in the type theory. The result of the discussion of hyperdoctrines could have gone otherwise. The notion of identity could have been constrained by definitional equality, which would object to the distinction made in MLTT between propositional and definitional identity. Or perhaps we would have found that identity admits only one proof, even were other predicates to admit of many, showing proof-relevant identity to have the UIP (uniqueness of identity proofs) property that we do not want. So our thesis was falsifiable and because the moderately neutral characterization of identity states that identity is inductively generated by reflexivity, we arrive at a compelling case that Path Induction (the J rule) is substantively supported.

Clearly these two reasons, independence and neutrality, are very closely related. To pull them apart slightly, I suggest we view the independence as an advantage in mathematical justifications and the neutrality as an advantage in philosophical justification. I don't want to claim this is a hard and fast distinction, but it does indicate an important way that I view this paper to have proceeded. The mathematical justification of the identity rules was a demonstration that the rules were admissible. To do this, we ensure the rules are continuous or conservative with the mathematical 'fragment' without the rules. This is similar to how conservative extensions are shown. Independence is this mathematical condition for showing admissibility. Neutrality, on the other hand, requires that the philosophical justification not assume the conclusion it aims to support. Were the analysis of hyperdoctrines replaced by a model of Martin-Löf Type Theory, we would have failed to be neutral. The question would have been decided before we had begun. After all, the model of MLTT will satisfy the identity elimination rule *by construction*, which would invalidate its philosophical import. We can imagine that giving a model for a system is all the justification a mathematician needs in order to feel comfortable with the rules. It seems that giving a model of HoTT does imply that HoTT is usable in mathematics (to some extent) and also that it is relatively consistent. But it does not satisfy the philosopher with respect to the concepts involved, especially one looking for a foundations of mathematics.

The characterization of identity in the last chapter is not new. Lawvere gives this adjoint characterization in [79] and this is why I call it "Lawvere's Law". But the conceptual conclusion that identity is the initial reflexive relation, the limit of reflexive relations, and inductively generated by reflexivity have all been said before. Peter Dybjer states that Lawvere's formulation "is similar to saying that 'equality is inductively generated by the reflexivity rule'" in his history of the identity type [38, slide 18]. Dan Licata suggests that "the real force of J is to assert" that identifications (paths) from a point $x : A$ are 'inductively defined' with generator $(x, \mathtt{refl}_x)$.[1] In addition, the proof assistant `Coq` explicitly defines the identity type as the inductively family generated by reflexivity.[2] But I wanted to establish a perhaps more accessible 'if and only if' relation between reflexivity and identity which led to Lawvere's Law. I could have also framed it in terms of categorical limits. After all, the adjunction is equivalent to saying that anything that implies a reflexive relation also implies the equality relation. Lawvere's Law would then look something like the following

$$Eq :\equiv \varinjlim_{R(x,x)} R(x, y)$$

which is equivalent to our statement with adjunctions. We can see this by a restatement. Let R be a reflexive relation, that is implied by Eq, then for any Q that implies R, there exists a unique arrow $Q \to Eq$ such that the following diagram commutes:



All this is to say that the conception derived in the previous chapter is well-established and widely used. The universality of the adjunction or limit gives us the harmony component and although this is not a definitional η rule, it is sufficient for harmony.

[1] See his post on the HoTT site http://homotopytypetheory.org/2011/04/10/just-kidding-understanding-identity-elimination-in-homotopy-type-theory/

[2] The single line of code is indicative of this:

```
Inductive id {A} : A -> A -> Type := idpath : forall x, id x x
```

Bibliography

[1] Andrew Arana. "Proof Theory in Philosophy of Mathematics". In: *Philosophy Compass* 5.4 (2010), pp. 336–347. DOI: 10.1111/j.1747-9991.2010.00282.x.

[2] Steve Awodey. *Category theory*. Clarendon Press, 2006. DOI: 10.1111/1467-9973.00225.

[3] Steve Awodey. "Structuralism, Invariance, and Univalence". In: *Philosophia Mathematica* 22.1 (2014), pp. 1–11. DOI: 10.1093/phimat/nkt030.

[4] Steve Awodey. "Type theory and homotopy". In: *Epistemology versus Ontology*. Vol. 27. Springer Netherlands, 2012. Chap. Chapter 9, pp. 183–201. DOI: 10.1007/978-94-007-4435-6_9.

[5] Steve Awodey and Andrej Bauer. "Propositions as [Types]". In: *Journal of Logic and Computation* 14.2 (2004), pp. 447–471.

[6] Steve Awodey, Pieter Hofstra, and Michael A. Warren. "Martin-Löf Complexes". In: *Annals of Pure and Applied Logic* 164.10 (2013), pp. 928–956. arXiv: 0906.4521v1.

[7] Steve Awodey, Álvaro Pelayo, and Michael A. Warren. "Voevodsky's Univalence Axiom in Homotopy Type Theory". In: *Notices of the American Mathematical Society* 60.9 (2013), p. 1164. DOI: 10.1090/noti1043.

[8] Steve Awodey and Michael A. Warren. "Homotopy theoretic models of identity types". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 146.1 (2009), pp. 45–55. DOI: 10.1017/S0305004108001783. arXiv: 0709.0248.

[9] John L. Bell. "Category theory and the foundations of mathematics". In: *British Journal for the Philosophy of Science* 32.4 (1981), pp. 349–358.

[10] Gianluigi Bellin et al. "Categorical proof theory of classical propositional calculus". In: *Theoretical Computer Science* 364.2 (2006), pp. 146–165. DOI: 10.1016/j.tcs.2006.08.002.

[11] Nuel D. Belnap. "Tonk, plonk and plink". In: *Analysis* 22.6 (1962), pp. 130–134.

[12] Jean Bénabou. "Introduction to bicategories". In: *Reports of the Midwest Category Seminar*. Springer. 1967, pp. 1–77.

[13] Paul Benacerraf. "Mathematical truth". In: *The Journal of Philosophy* 70.19 (1973), pp. 661–679.

[14] Paul Benacerraf. "What numbers could not be". In: *The Philosophical Review* 74.1 (1965), pp. 47–73.

[15] Paul Benacerraf and Hilary Putnam, eds. *Philosophy of mathematics: Selected readings*. Cambridge University Press, 1964.

[16] Gustav Bergmann. "Logical Positivism, Language, and the Reconstruction of Metaphysics". In: *The Linguistic Turn: Essays in Philosophical Method*. Ed. by Richard Rorty. The University of Chicago Press, 1992. Chap. 3, pp. 63–71.

[17] Marc Bezem, Thierry Coquand, and Simon Huber. "A model of type theory in cubical sets". In: *19th International Conference on Types for Proofs and Programs (TYPES 2013)*. Vol. 26. 2014, pp. 107–128.

[18] Paul A. Boghossian. "Inferential role semantics and the analytic/synthetic distinction". In: *Philosophical Studies* 73.2 (1994), pp. 109–122. DOI: 10.1007/BF01207660.

[19] Robert Brandom. *Articulating Reasons: An Introduction to Inferentialism*. Harvard University Press, 2001, p. 240. DOI: 10.1111/j.1468-0149.2004.0335a.x.

[20] Robert Brandom. *From Empiricism to Expressivism: Brandom Reads Sellars*. Harvard University Press, 2014.

[21] Robert Brandom. *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press, 1994. DOI: 10.2307/2940874.

[22] James Robert Brown. *Philosophy of mathematics: A contemporary introduction to the world of proofs and pictures*. 2008.

[23] Rudolf Carnap. *Logical Syntax of Language*. Philosophy of mind and language. Routledge, 2001.

[24] Rudolf Carnap and Paul Arthur Schilpp. *The Philosophy of Rudolf Carnap*. 1963.

[25] Lewis Carroll. "What the tortoise said to Achilles". In: *Mind* 4.14 (1895), pp. 278–280.

[26] Ian Chiswell and Wilfrid Hodges. *Mathematical Logic*. Oxford University Press, 2007.

[27] Thierry Coquand. "Equality and dependent type theory". In: *Bologna lectures* (2011).

[28] Thierry Coquand. *Univalent Foundation of Mathematics*. 2014.

[29] Tierry Coquand. "A remark on singleton types". 2014.

[30] Leo Corry. *Modern algebra and the rise of mathematical structures*. 2004.

[31] Roy L Crole. *Categories for types*. Cambridge University Press, 1993.

[32] Kosta Došen and Zoran Petrić. *Proof-Theoretical Coherence*. King's College Publications, 2004.

[33] Michael Dummett. *Elements of intuitionism*. Oxford University Press, 2000.

[34] Michael Dummett. *The logical basis of metaphysics*. Vol. 5. Harvard University Press, 1991.

[35] Michael Dummett. *Truth and Other Enigmas*. Harvard University Press, 1978.

[36] Michael Dummett. "What I Know When I Know a Language". In: *Seas of Language*. Oxford: Clarendon Press, 1996. Chap. 3, pp. –19. DOI: 10.1093/0198236212.003.0003.

[37] Peter Dybjer. "Program Testing and the Meaning Explanations of Intuitionistic Type Theory". In: *Epistemology versus Ontology*. Springer, 2012, pp. 215–241. DOI: 10.1007/978-94-007-4435-6_11.

[38] Peter Dybjer. *(What I Know About) the History of the Identity Type*. Workshop on Identity Types, Uppsala. Nov. 2006. URL: http://www.cse.chalmers.se/~peterd/papers/historyidentitytype.pdf.

[39] Peter Dybjer et al., eds. *Epistemology versus Ontology: Essays on the Philosophy of Foundations of Mathematics in Honour of Per Martin-Löf*. Springer Berlin Heidelberg, 2012.

[40] Solomon Feferman. "Categorical foundations and foundations of category theory". In: *Logic, Foundations of Mathematics and Computability Theory*. Ed. by Robert E. Butts and Jaakko Hintikka. Springer, 1977, pp. 149–169. DOI: 10.1007/978-94-010-1138-9_9.

[41] Solomon Feferman. "Foundations of unlimited category theory: What remains to be done". In: *The Review of Symbolic Logic* 6.01 (2013), pp. 6–15.

[42] Solomon Feferman. "What rests on what? The proof-theoretic analysis of mathematics". In: *15th International Wittgenstein Symposium: Philosophy of Mathematics* (1992), pp. 16–23.

[43] Hartry Field. *Science without numbers*. Princeton University Press, 1980.

[44] Nissim Francez and Gilad Ben-Avi. "Proof-Theoretic Semantic Values for Logical Operators". In: *The Review of Symbolic Logic* 4.03 (Sept. 2011), pp. 466–478. DOI: 10.1017/S1755020311000098.

[45] Nissim Francez and Roy Dyckhoff. "Proof-theoretic semantics for a natural language fragment". In: *Linguistics and Philosophy* 33.6 (May 2010), pp. 447–477. DOI: 10.1007/s10988-011-9088-3.

[46] Nissim Francez, Roy Dyckhoff, and Gilad Ben-Avi. "Proof-Theoretic Semantics for Subsentential Phrases". In: *Studia Logica* 94.3 (2010), pp. 381–401. DOI: 10.1007/s11225-010-9241-y.

[47] Gottlob Frege. "Sense and Reference". In: *The Philosophical Review* 57.3 (1948), pp. 209–230.

[48] Gottlob Frege. *The Foundations of Arithmetic: A Logico-mathematical enquiry into the concept of number*. Ed. by J. L. Austin. Second Rev. New York: Harper and Brothers, 1960.

[49] Gerhard Gentzen. "Investigations into logical deduction". In: *American philosophical quarterly* 1.4 (1964), pp. 288–306. DOI: 10.2307/20009142.

[50] Gerhard Gentzen. "Untersuchungen über das logische Schließen. I". In: *Mathematische Zeitschrift* 39.1 (1935), pp. 176–210. DOI: 10.1007/BF01201353.

[51] Donald Gillies and Herbert Mehrtens. "Revolutions in Mathematics. Edited by Donald Gillies. Oxford (Clarendon Press). 1992. 353". In: 20 (1993), pp. 320–323.

[52] Jean-Yves Girard, Paul Taylor, and Yves Lafont. *Proofs and types*. 1989.

[53] Johan Georg Granström. *Treatise on Intuitionistic Type Theory*. Springer, 2011. DOI: 10.1007/978-94-007-1736-7.

[54] Mark Greenberg and Gilbert Harman. "Conceptual Role Semantics". In: *Notre Dame Journal of Formal Logic* 23.2 (1982), pp. 242–256. DOI: doi:10.1305/ndjfl/1093883628.

[55] Victor Harnik and Michael Makkai. "Lambek's categorical proof theory and Läuchli's abstract realizability". In: *The Journal of Symbolic Logic* 57.1 (1992), pp. 200–230.

[56] Geoffrey Hellman. "Does Category Theory Provide a Framework for Mathematical Structuralism?" In: *Philosophia Mathematica* 11.3 (2003), pp. 129–157. DOI: 10.1093/philmat/11.2.129.

[57] Martin Hofmann. "Elimination of extensionality in Martin-Löf type theory". In: *Types for proofs and programs* (1994), pp. 166–190. DOI: 10.1007/3-540-58085-9_76.

[58] Martin Hofmann and Thomas Streicher. "The groupoid interpretation of type theory". In: *Twenty Five Years of Constructive Type Theory*. Ed. by Giovanni Sambin and Jan M. Smith. Oxford University Press, 1998. Chap. 1, pp. 1–29.

[59] Martin Hofmann and Thomas Streicher. "The groupoid model refutes uniqueness of identity proofs". In: *Proceedings Ninth Annual IEEE Symposium on Logic in Computer Science* (1994). DOI: 10.1109/lics.1994.316071.

[60] John Horty. *Frege on Definitions*. Oxford University Press, 2007.

[61] William A. Howard. "The formulae-as-types notion of construction". In: *The Curry-Howard Isomorphism*. Ed. by Philippe De Groote. Academia, 1995, pp. 479–490.

[62] Bart Jacobs. *Categorical Logic and Type Theory*. Elsevier, 1999, p. 784.

[63] Andre Joyal. *Notes on quasi-categories*. 2008.

[64] Fairouz Kamareddine, Twan Laan, and Rob Nederpelt. *A Modern Perspective on Type Theory: Form its Origins until Today*. Kluwar Academic Publishers, 2005.

[65] Peter Knijnenburg and Frank Nordemann. "Partial hyperdoctrines: categorical models for partial function logic and Hoare logic". In: *Mathematical Structures in Computer Science* 4.2 (1994), p. 117. DOI: 10.1017/S0960129500000414.

[66] Arnold Koslow. *A structuralist theory of logic*. Cambridge University Press, 2005.

[67] Ralf Kromer. *Tool and Object: A History and Philosophy of Category Theory*. Birkhäuser, 2007. DOI: 10.1007/978-3-7643-7524-9.

[68] James Ladyman and Stuart Presnell. "A Primer on Homotopy Type Theory". 2014.

[69] James Ladyman and Stuart Presnell. "Does Homotopy Type Theory Provide a Foundation for Mathematics". Nov. 2014.

[70] James Ladyman and Stuart Presnell. "Identity in Homotopy Type Theory, Part I : The Justification of Path Induction". Nov. 2014.

[71] Joachim Lambek. "Deductive systems and categories II. Standard constructions and closed categories". In: *Category theory, homology theory and their applications I* (1969), pp. 76–122. DOI: 10.1007/BFb0079385.

[72] Joachim Lambek and Philip J Scott. *Introduction to higher-order categorical logic*. Cambridge University Press, 1986.

[73] Mark Norris Lance. "Some Reflections on the Sport of Language". In: *Noûs* 32.S12 (1998), pp. 219–240. DOI: 10.1111/0029-4624.32.s12.10.

[74] Mark Norris Lance and John O'Leary-Hawthorne. *The Grammar of Meaning: Normativity and Semantic Discourse*. Cambridge University Press, 2008.

[75] Elaine Landry. "Category theory: the language of mathematics". In: *Philosophy of Science* 66.S1 (1999), S14. DOI: 10.1086/392712.

[76] Elaine Marie Landry. "Category-Theoretic Realism: A Linguistic approach to the philosophy of mathematics". PhD thesis. University of Western Ontario, 1997.

[77] F. William Lawvere. "Adjointness in foundations". In: *Dialectica* 23.3-4 (1969), pp. 281–296. DOI: 10.1111/j.1746-8361.1969.tb01194.x.

[78] F. William Lawvere. "An Elementary Theory of the Category of Sets (Long Version) with Commentary". In: *Reprints in Theory and Applications of Categories* 11.11 (2005), pp. 1–35. DOI: 10.1073/pnas.52.6.1506.

[79] F. William Lawvere. "Equality in hyperdoctrines and comprehension schema as an adjoint functor". In: *Applications of Categorical Algebra* (1970).

[80] F. William Lawvere. "Some thoughts on the future of category theory". In: *Category Theory*. Vol. Lecture No. Springer Berlin Heidelberg, 1991, pp. 1–13. DOI: 10.1007/BFb0084208.

[81] Øystein Linnebo and Richard Pettigrew. "Category Theory as an Autonomous Foundation". In: *Philosophia Mathematica* 19.3 (Aug. 2011), pp. 227–254. DOI: 10.1093/phimat/nkr024.

[82] Saunders Mac Lane and Ieke Moerdijk. *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Springer, 1992.

[83] Michael Makkai. "The fibrational formulation of intuitionistic predicate logic I: completeness according to Gödel, Kripke, and Läuchli. Part 1". In: *Notre Dame Journal of Formal Logic* 34.3 (1993), pp. 334–377. DOI: 10.1305/ndjfl/1093634727.

[84] Lorenzo Malatesta. "Some proposals for the set-theoretic foundations of category theory". In: *Rivista Italiana di Filosofia Analitica Junior* 2.2 (2011), pp. 41–58.

[85] Jean-Pierre Marquis. "Category Theory". In: *Stanford Encyclopedia of Philosophy* (2013).

[86] Jean-Pierre Marquis. "Category theory and the foundations of mathematics: philosophical excavations". In: *Synthese* 103.3 (1995), pp. 421–447. DOI: 10.1007/BF01089735.

[87] Jean-Pierre Marquis. *From a geometrical point of view: a study of the history and philosophy of category theory*. Springer, 2008. DOI: 10.1007/978-1-4020-9384-5.

[88] David Marshall. "The Implications of Robert Brandom's Inferentialism for Intellectual History". In: *History and Theory* 52.February (2013), pp. 1–31.

[89] Per Martin-Löf. "Constructive Mathematics and Computer Programming". In: *Logic, Methodology and Philosophy of Science VI, Proceedings of the Sixth International Congress of Logic, Methodology and Philosophy of Science*. Vol. 104. Studies in Logic and the Foundations of Mathematics 104. Elsevier, 1982, pp. 153–175. DOI: 10.1016/S0049-237X(09)70189-2.

[90] Per Martin-Löf. "Hauptsatz for the intuitionistic theory of iterated inductive definitions". In: *Studies in Logic and the Foundations of Mathematics* 63 (1971), pp. 179–216.

[91] Per Martin-Löf. *Intuitionistic Type Theory*. Bibliopolis, 1984.

[92] Per Martin-Löf. "On the meanings of the logical constants and the justifications of the logical laws". In: *Nordic Journal of Philosophical Logic* 1.1 (1996), pp. 11–60.

[93] Per Martin-Löf. "Truth of a proposition, evidence of a judgement, validity of a proof". In: *Synthese* 73.3 (1987), pp. 407–420.

[94] John Mayberry. "What is Required of a Foundation for Mathematics?" In: *Philosophia Mathematica* 2.1 (1994), pp. 16–35. DOI: 10.1093/philmat/2.1.16.

[95] Colin McGinn. *Logical Properties: Identity, Existence, Predication, Necessity, Truth*. Oxford University Press, 2000. DOI: 10.1215/00318108-111-3-462.

[96] Colin McLarty. "Categorical Foundations and Mathematical Practice". In: *Philosophia Mathematica* 20.1 (2012), pp. 111–113. DOI: 10.1093/phimat/nkr041.

[97] Colin McLarty. "Exploring Categorical Structuralism". In: *Philosophia Mathematica* 12.1 (Feb. 2004), pp. 37–53. DOI: 10.1093/philmat/12.1.37.

[98] Colin McLarty. "Learning from questions on categorical foundations". In: *Philosophia Mathematica* 13.1 (2005), pp. 44–60. DOI: 10.1093/philmat/nki006.

[99] Grigori Mints. *A Short Introduction to Intuitionistic Logic*. Kluwar Academic Publishers, 2000.

[100] Danièle Moyal-Sharrock. "Understanding Wittgenstein's On Certainty". In: (Apr. 2004). DOI: 10.1057/9780230504462.

[101] Julien Murzi and Ole Thomassen Hjortland. "Inferentialism and the categoricity problem: reply to Raatikainen". In: *Analysis* 69.3 (June 2009), pp. 480–488. DOI: 10.1093/analys/anp071.

[102] B Nordström, K Petersson, and J.M. M Smith. "Martin-Löf's Type Theory". In: *Handbook of Logic in Computer Science*. Vol. 5. October. 2000, pp. 1–37.

[103] Peter Pagin. *Compositionality, understanding, and proofs*. Aug. 2009. DOI: 10.1093/mind/fzp093.

[104] Valeria de Paiva. "Categorical Proof Theory and Linear Logic". In: (1996).

[105] Christine Paulin-Mohring. "Inductive definitions in the system Coq rules and properties". English. In: *Typed Lambda Calculi and Applications*. Ed. by Marc Bezem and JanFriso Groote. Vol. 664. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1993, pp. 328–345. DOI: 10.1007/BFb0037116.

[106] Dusko Pavlović. "Maps II: Chasing diagrams in categorical proof theory". In: *Logic Journal of IGPL* 4.2 (1996), pp. 159–194.

[107] Christopher Peacocke. "The Metaphysics of Concepts". In: *Mind* 100.October 1991 (1991).

[108] Jaroslav Peregrin. "Inferentialism and the Compositionality of Meaning". In: *International Review of Pragmatics* 1.1 (Jan. 2009), pp. 154–181. DOI: 10.1163/187731009X455875.

[109] Jaroslav Peregrin. "What is inferentialism?" In: *Inference, Consequence, and Meaning: Perspectives on Inferentialism* (2012), pp. 3–16.

[110] Frank Pfenning and Rowan Davies. "A Judgemental Reconstruction of Modal Logic". In: *Mathematical Structures in Computer Science* 11.4 (Aug. 2001), pp. 511–540. DOI: 10.1017/S0960129501003322.

[111] Andrew Pitts. "Categorical logic". In: *Handbook of Logic in Computer Science*. Ed. by S. Abramsky, Dov M. Gabbay, and T.S.E. Maibaum. Vol. VI. May. Oxford University Press, 1995.

[112] Andrew Pitts. *Notes on Categorical Logic*. 1991.

[113] Dag Prawitz. "Meaning approached via Proofs". In: *Synthese* 148.3 (2006), pp. 507–524. DOI: 10.1007/s.

[114] Dag Prawitz. *Natural Deduction: a Proof-Theoretical Study*. Stockholm: Almqvist & Wiksell, 1965.

[115] Dag Prawitz. "On the Idea of a General Proof Theory". In: *Synthese* 27.1/2 (1974), pp. 63–77. DOI: 10.1556/Acr.8.2007.1.1.

[116] Arthur N. Prior. "The Runabout Inference-Ticket". In: *Analysis* 21.2 (1960), pp. 38–39.

[117] Hilary Putnam. "Philosophy of Logic". In: *Mathematics, matter, and method*. 1971. Chap. 28.

[118] W. V. O. Quine. "Speaking of Objects". In: *Proceedings and Addresses of The American Philosophical Association* 31 (1957), pp. 5–22.

[119] Willard Van Orman Quine. "On natural deduction". In: *The Journal of Symbolic Logic* 15.2 (1950), pp. 93–102.

[120] Willard Van Orman Quine. "Two dogmas of empiricism". In: *The Philosophical Review* 60.1 (1951), pp. 20–43.

[121] Stephen Read. "General elimination harmony and higher-level rules". In: *Dag Prawitz on Proofs and Meaning*. Ed. by Heinrich Wansing. Springer International Publishing, 2014. Chap. 13, pp. 293–312.

[122] Stephen Read. "Identity and harmony". In: *Analysis* 64.2 (2004), pp. 113–119. DOI: 10.1111/j.1467-8284.2004.00469.x.

[123] Egbert Rijke. "Homotopy Type Theory". M.S. Thesis. Utrecht University, 2012.

[124] Andrei Rodin. "Categories without structures". In: *Philosophia Mathematica* 19.1 (2011), pp. 20–46. DOI: 10.1093/philmat/nkq027. arXiv: 0907.5143.

[125] Richard Rorty. *Philosophy and the mirror of nature*. 1979. DOI: 10.2307/2107303.

[126] Richard Rorty. *The Linguistic Turn: Essays in Philosophical Method*. University of Chicago Press, 1992, p. 407. DOI: 10.1111/j.1468-0149.1968.tb00368.x.

[127] Bertrand Russell. *A Critical Exposition of the Philosophy of Leibniz*. Routledge, 2005.

[128] Bertrand Russell. "Mathematical logic as based on the theory of types". In: *American Journal of Mathematics* 30.3 (1908), pp. 222–262.

[129] Georg Schiemer. "Invariants and mathematical structuralism". In: *Philosophia Mathematica* 22.1 (2014), pp. 70–107. DOI: 10.1093/philmat/nkt032.

[130] Peter Schroeder-Heister. "Proof-Theoretic Semantics". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N Zalta. Summer 201. 2014.

[131] Peter Schroeder-Heister. "Proof-Theoretic Semantics, Self-Contradiction, and the Format of Deductive Reasoning". In: *Topoi* 31.1 (Mar. 2012), pp. 77–85. DOI: 10.1007/s11245-012-9119-x.

[132] Peter Schroeder-Heister. "Proof-theoretic versus model-theoretic consequence". In: *The Logica Yearbook* (2007), pp. 1–12.

[133] Peter Schroeder-Heister. "Validity concepts in proof-theoretic semantics". In: *Synthese* 148.3 (2006), pp. 525–571. DOI: 10.1007/s11229-004-6296-1.

[134] John R. Searle. *Minds, Brains and Science*. Cambridge, Massachusetts: Harvard University Press, 1984.

[135] Robert A. G. Seely. "Hyperdoctrines, Natural Deduction and the Beck Condition". In: *Mathematical Logic Quarterly* 29 (1983), pp. 505–542. DOI: 10.1002/malq.19830291005.

[136] Wilfrid Sellars. "Empiricism and the Philosophy of Mind". In: *Minnesota studies in the philosophy of science*. Vol. 1. 1956, pp. 253–329.

[137] Wilfrid Sellars. "Inference and Meaning". In: *Mind* 62.247 (1953), pp. 313–338.

[138] Wilfrid Sellars. "Language, Rules, and Behavior". In: *John Dewey: Philosophy of Science and Freedom*. Ed. by Sidney Hook. New York: Dial Press, 1949, pp. 289–315.

[139] Wilfrid Sellars. "Meaning as functional classification". In: *Synthese* 3.3-4 (1974), pp. 417–437. DOI: 10.1007/BF00484606.

[140] Wilfrid Sellars. "Some Reflections on Language Games". In: *Philosophy of Science* 21.3 (1954), pp. 204–228.

[141] Morten Heine Sørensen and Pawel Urzyczyn, eds. *Lectures on the Curry-Howard isomorphism*. Vol. 149. Elsevier, 2006.

[142] Florian Steinberger. "Harmony and Logical Inferentialism". PhD thesis. University of Cambridge, 2009.

[143] Florian Steinberger. "Not so stable". In: *Analysis* 69.4 (July 2009), pp. 655–661. DOI: 10.1093/analys/anp100.

[144] Florian Steinberger. "What Harmony Could and Could Not Be". In: *Australasian Journal of Philosophy* 89.4 (Dec. 2011), pp. 617–639. DOI: 10.1080/00048402.2010.528781.

[145] J. T. Stevenson. "Roundabout the Runabout Inference-Ticket". In: *Analysis* 21.6 (1961), pp. 124–128.

[146] Manfred Egon Szabo. *The collected papers of Gerhard Gentzen*. Vol. 160. North-Holland Amsterdam, 1969.

[147] Neil Tennant. "Existence and Identity in Free Logic: A Problem for Inferentialism?" In: *Mind* 116.464 (Nov. 2007), pp. 1055–1078. DOI: 10.1093/mind/fzm1055.

[148] Neil Tennant. "Inferential Semantics for First-Order Logic: Motivating Rules of Inference from Rules of Evaluation". In: *The Force of Argument: Essays in Honor of Timothy Smiley* (2009), pp. 223–257.

[149] Neil Tennant. "Inferentialism, Logicism, Harmony, and a Counterpoint". In: *Essays for Crispin Wright: Logic, language, and mathematics* 2 (2007), pp. 105–132.

[150] Neil Tennant. "Intentionality, Syntactic Structure and the Evolution of Language". In: *Minds, Machines, and Evolution: Philosophical Studies*. Ed. by Christopher Hookway. Cambridge University Press, 1984.

[151] Neil Tennant. "Rule-Circularity and the Justification of Deduction". In: *The Philosophical Quarterly* 55.221 (2005), pp. 625–648. DOI: 10.1111/j.0031-8094.2005.00420.x. eprint: http://pq.oxfordjournals.org/content/55/221/625.full.pdf+html.

[152] A. S. Troelstra. "Aspects of constructive mathematics". In: *Handbook of Mathematical Logic*. Ed. by Jon Barwise. North-Holland Publishing Company, 1977. Chap. D5, pp. 973–1052.

[153] A. S. Troelstra and Dirk van Dalen. *Constructivism in Mathematics: An introduction*. Vol. I. Elsevier Science Publishers, 1988.

[154] A. S. Troelstra and H. Schwichtenberg. *Basic proof theory*. 2000.

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

[156] Heinrich Wansing. "The idea of a proof-theoretic semantics and the meaning of logical operators". In: *Studia Logica* 64.1 (2000), pp. 3–20.

[157] Eugene P Wigner. "The unreasonable effectiveness of mathematics in the natural sciences". In: *Communications on pure and applied mathematics* 13.1 (1960), pp. 1–14.

[158] C. J. F. Williams. *What is identity?* Oxford University Press, 1989.

[159] Andrzej Wisniewski. "Questions and inferences". In: *Logique & Analyse* 173-174-175 (2001), pp. 5–43.

[160] Ludwig Wittgenstein. *On Certainty*. Ed. by G. E. M. Anscombe, G. H. von Wright, and Denis Paul. Oxford: Basil Blackwell Ltd, 1969.

[161] Ludwig Wittgenstein. *Philosophical investigations*. Ed. by G.E.M. Anscombe. Second. John Wiley & Sons, 1958.

[162] Ludwig Wittgenstein. *Tractatus Logico-Philosophicus (Trans. Pears and McGuinness)*. Routledge, 1994, pp. 332–335.

[163] Crispin Wright. *Wittgenstein on the Foundations of Mathematics*. Gregg Revivals, 1980.