UNIVALENCE, FOUNDATIONS
AND PHILOSOPHY
With a Sheaf-Shaped Appendix

DIMITRIOS TSEMENTZIS

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Philosophy

Advisers: Hans Halvorson & John Burgess

SEPTEMBER 2016
Abstract

The Univalent Foundations (UF) of mathematics provide a foundation for mathematics entirely independent from Cantorian set theory. This development raises important questions: In what sense is UF a new foundation? How does it relate to set theory? How can it be justified philosophically? It also raises fundamental methodological questions about analytic philosophy: how are we to justify the pervasive use of first-order logic and set theory when confronted with a foundation of mathematics in which neither plays an essential role? This dissertation aims to answer all these questions.

In Chapter 1, I orient my project by investigating the relation between philosophy, the foundations of mathematics and formal logic. Then, in Chapter 2 I argue that UF is better-able to live up to the ideal of a structuralist foundation than other proposals and respond to several challenges against the foundational aspirations of UF.

In the next two chapters I compare UF to other foundational proposals. In Chapter 3 I argue for a pluralistic picture between UF and ZFC, examine the extent to which Homotopy Type Theory can receive a pre-formal “meaning explanation” independent of set theory and respond to a potent objection raised by Hellman and Shapiro against non-set-theoretic foundations of mathematics. In Chapter 4 I examine alternative structuralist foundations and argue that Makkai’s Type-Theoretic Categorical Foundations of Mathematics (TTCFM) emerges as the most serious contender to UF. I then compare UF and TTCFM on several fronts, including on their intended semantics (∞-groupoids vs. ∞-categories), offering an argument in favour of ∞-groupoids as the basic objects of a structuralist foundation.

In the final chapter I develop a mathematical logic (“n-logic”) for UF by extending Makkai’s system of First-Order Logic with Dependent Sorts (FOLDS). I define the syntax and proof system for n-logic, prove soundness with respect to both homotopy-theoretic
and set-theoretic semantics, and sketch some applications. This establishes a mathematical logic for UF that provides the groundwork for a new kind of formal philosophy. And after that comes the time, in the evening light, to dance...
Acknowledgments

I would like to thank (in alphabetical order) Benedikt Ahrens, Steve Awodey, John Burgess, Dan Grayson, Hans Halvorson, Des Hogan, Chris Kapulkin, Per Martin-Löf, Peter Lumsdaine, Anders Mörtberg, Colin McLarty, Paige North, Alexander Nehamas, Mike Shulman, Tim Stoll, Nat Tabris, Vladimir Voevodsky, Dan Wolt and Fernando Zalamea for helpful conversations and correspondence as well as inspiration and confrontation during the preparation of this work.
## Contents

Abstract iii  
Acknowledgments v  
List of Figures x  

Chapter 1. Foundations and (Analytic) Philosophy 1  
1.1. Old Foundations and Analytic Philosophy 2  
1.2. A New Kind of Synthetic Geometry 8  
1.3. Univalent Foundations and Philosophy 15  
1.4. A Challenge for Recent Metaphysics 22  
1.5. Beyond Analytic Philosophy 29  
Appendix in the Shape of a Small Spiral 36  

Chapter 2. Univalent Foundations as Structuralist Foundations 71  
2.1. Univalent Foundations, HoTT and intensional MLTT 74  
2.2. Structuralist Foundations of Mathematics 79  
2.2.1. ZFC and ETCS are not structuralist foundations 83  
2.2.2. From Ontology to Language 92  
2.3. Univalence and (SFOM) 96  
2.4. Criteria of Identity as Homotopy Equivalence 100  
2.4.1. Component-wise isomorphism as a criterion of identity 101  
2.4.2. The general method for (CI) 105  
2.5. Formalizing all of mathematics in HoTT 111  
2.5.1. Is HoTT a “big-f” Foundation for all of mathematics? 111
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5.1.</td>
<td>Pre-formal Groupoids</td>
<td>232</td>
</tr>
<tr>
<td>4.5.2.</td>
<td>Pre-formal Categories</td>
<td>233</td>
</tr>
<tr>
<td>4.5.3.</td>
<td>Bishop or Frege?</td>
<td>235</td>
</tr>
<tr>
<td>4.5.4.</td>
<td>Next-level sets are groupoids</td>
<td>237</td>
</tr>
<tr>
<td>Appendix</td>
<td>Homotopy equivalence as FOLDS equivalence</td>
<td>241</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Homotopy Model Theory: A Mathematical Logic for UF</td>
<td>247</td>
</tr>
<tr>
<td>5.1.</td>
<td>Preliminaries</td>
<td>251</td>
</tr>
<tr>
<td>5.2.</td>
<td>Syntax of $n$-logic</td>
<td>258</td>
</tr>
<tr>
<td>5.3.</td>
<td>Homotopy Semantics of $n$-logic</td>
<td>263</td>
</tr>
<tr>
<td>5.4.</td>
<td>Set-Theoretic semantics for 1-logic</td>
<td>272</td>
</tr>
<tr>
<td>5.5.</td>
<td>Proof System for $n$-logic</td>
<td>282</td>
</tr>
<tr>
<td>5.6.</td>
<td>Soundness</td>
<td>289</td>
</tr>
<tr>
<td>5.7.</td>
<td>Examples and Applications</td>
<td>297</td>
</tr>
<tr>
<td>5.8.</td>
<td>Prospects</td>
<td>307</td>
</tr>
<tr>
<td>Appendix</td>
<td>in the Shape of a Sheaf</td>
<td>313</td>
</tr>
<tr>
<td>Bibliography</td>
<td></td>
<td>337</td>
</tr>
</tbody>
</table>
List of Figures

1. A sparse spiral ................................................................. 36
2. A dense spiral ................................................................. 37
3. Sphere and Donut ............................................................... 67
4. Ulam’s spiral grid ............................................................. 68
5. Ulam’s prime diagonals ..................................................... 69
6. A shape observable from another shape ................................. 157
7. A portion visible from a point of a shape ................................. 158
8. Spatial $Id$-elimination .................................................... 171
CHAPTER 1

Foundations and (Analytic) Philosophy

There are two main reasons one may wish to introduce new formal methods and results in philosophy. On the one hand, one may do so with the hope of substantiating a philosophical thesis; for example, by taking Kripke semantics for modal logic to lend support to a possible-worlds understanding of modality. On the other hand, one may study some new formalism as a kind of entryway into new forms and styles of thinking, without a specific thought as to their direct applicability to philosophical disputes; for example, this was arguably the spirit in which students of Plato’s Academy were encouraged to study geometry, i.e. not in order to approach philosophy by drawing shapes and carrying out demonstrations, but by letting the geometric way of thinking sanitize and prepare the mind in some way or another. I want to be clear about this from the outset: in this entire dissertation, no formal apparatus is introduced with the purpose of substantiating some philosophical thesis or in the hopes of settling some pre-existing dispute. Rather, the introduction of this formal apparatus is made entirely with the intention of exploring new forms and styles of thinking.

Since these few sketchy remarks can hardly suffice, the purpose of this Chapter is to make precise the relation between philosophy and the foundations of mathematics as I see it, as a prelude and at the same time a justification for all the immense and numbing pedantry that follows, before the final dance in the Appendix. In Section 1.1 I explain the link between the “old foundations” and analytic philosophy, by examining exactly what Russell thought the connection could be between his formal work in mathematical logic and its application to philosophical problems. Then, in Section 1.2, I present a strain
of thinking in synthetic geometry that originates at around the same time as Cantor and Frege were laying the groundwork for the “old foundations” and connect it to a ubiquitous methodology in contemporary mathematics associated with Grothendieck. In Section 1.3 I explain how recent developments in the foundations of mathematics connect to philosophy in a way very similar to the one Russell envisioned connected his own work on foundations to analytic philosophy. Given this link, I explain in Section 1.4 how this situation poses a dilemma for parts of analytic metaphysics that depend essentially on set-theoretic intuitions and I elaborate on a specific form of this dilemma pertinent to recent literature on “grounding”. Finally, in Section 1.5, I compare my criticism of analytic philosophy to some alternative recent criticisms and examine the methodological difficulty raised by the fact that my criticism is located inside a thesis of analytic philosophy. This sets the stage for an alternative version of this Chapter, given as an Appendix. Thus, the relation between Sections 1.1-1.5 of Chapter 1 and its “spiral-shaped” Appendix serves as the methodological blueprint for the relation between the whole dissertation and its “sheaf-shaped” Appendix.

1.1. Old Foundations and Analytic Philosophy

The dominant narrative in the foundations of mathematics goes roughly as follows: the increasing complexity of mathematical analysis in the 19th century made it imperative to reconstrue all mathematical propositions as logical (i.e. analytic) truths about simple entities – indeed, ideally, as logic itself. And this demand – through Cantor, Frege, Russell, Dedekind, Hilbert, Peano, Zermelo, Gödel and many others – eventually gave rise to set theory and first-order logic as we now understand it together with an all-encompassing foundation of mathematics in the form of Zermelo-Fraenkel set theory (ZF(C)). Surely it is no exaggeration to say that the fact that all of mathematics could conceivably be encoded in terms of a single binary predicate (governed by a finitely
specifiable list of axioms) was one of the major discoveries of the 20th century, in any field of knowledge. This monumental achievement, I want to claim, also played a key role in bringing about the school of philosophy that has come to be known as *analytic philosophy*.

The story I want to tell about how this came about goes as follows. In the First Critique’s Doctrine of Method (especially [A713/B742]) Kant provides the clearest exposition of his view on mathematics in his critical period; he explains that mathematics proceeds by *constructing* intuitions adequate to *a priori* concepts, not by *analyzing* such concepts (which is the task of philosophy). Russell’s very influential criticism of Kant was that a strong enough logic (e.g. first-order predicate logic) could *compensate* for this element of “construction” without invoking some kind of pure intuition provided by the human subject. Russell concludes that Kant’s doctrine of construction in pure intuition was metaphysical froth added to the mix in order to *compensate* for the deficient (Aristotelian) logic that Kant was working with. This point of view came to be known as the “compensation thesis”. Eventually, at least in the Anglo-Saxon tradition, the compensation thesis comes to be accepted as a more or less definitive objection against Kant’s conception of mathematics. As a result, Kantian views on the synthetic (and *a priori*) nature of mathematics are more or less dismissed as the inevitable outgrowth of a much too impoverished logic.

1Russell maintained this criticism of Kant, in one form or another, throughout his life, cf. in particular [157] and [155], quoted below. As Friedman writes: “Russell [...] habitually blamed all the traditional obscurities surrounding space and geometry – including Kant’s views of course – on ignorance of the modern theory of relations and uncritical reliance on [Aristotelian logic].” ([47], p.457)

2For more discussion on Kant’s views on geometry see the Parsons-Hintikka debate [66,146] as well as Friedman’s very influential recasting of the compensation thesis in [47,48]. It should also be made clear that the issue of whether or not Kant thought that demonstrations themselves also involved pure intuition (rather than just the construction of the concepts that they were to be applied to) remains a topic of controversy. For the latest installment, see Hogan [69].

3The Neo-Kantians of the Marburg school did not share this conclusion, needless to say. In particular, Cassirer thought that “Russell’s logistic” not only failed to undermine the critical philosophy, but indeed could provide it exactly with the kind of raw material that it needs for its future development. For an illuminating analysis of the Neo-Kantian “absorption” of Russell and Frege’s work on the foundations of mathematics by Cassirer see Heis [59,60].
From there, it is a very natural step to take, to think that if a logic is powerful enough to provide a foundation for mathematics that does not rely on “metaphysical froth” then it should be powerful enough to be deployed to tackle other philosophical problems, beyond merely that of how mathematical propositions acquire their certainty. It is unclear whether Frege himself had such ambitions, but Russell undoubtedly did, and it is exactly this step that Russell took. What Russell had in mind is perhaps best revealed in the following passage of his programmatic [155], from which I will quote extensively:

The proof that all pure mathematics, including Geometry, is nothing but formal logic, is a fatal blow to the Kantian philosophy. [...] The whole doctrine of a priori intuitions, by which Kant explained the possibility of pure mathematics, is wholly inapplicable to mathematics in its present form. The Aristotelian doctrines of the schoolmen come nearer in spirit to the doctrines which modern mathematics inspire; but the schoolmen were hampered by the fact that their formal logic was

---

4 Even at his most philosophical, e.g. [45], it seems unclear to me whether Frege is writing with the purpose of applying symbolic logic to philosophical questions, rather than using philosophical questions to clarify the principles of his system. There is little doubt that Frege’s initial and overarching interest was in the foundations of mathematics (especially arithmetic) as is evident throughout the Grundlagen and also apparent from his own academic trajectory. Whether or not he thought of his foundational work as a way of attaining definitive solutions to philosophical problems is less clear, and in my opinion probably false. Perhaps Frege was restrained in this regard by the near-total acceptance of the Kantian system within German academia – as evidenced, for example, by the excessive (perhaps even ironic) caution he exercises when raising objections to Kantian doctrine:

In order not to open myself up to the criticism of carrying on a picayune search for faults in the work of a genius whom we look up to only with thankful awe [...] If Kant erred with respect to arithmetic, this does not detract essentially, I think, from his merit. (Grundlagen, §89, as translated in [20])

As this last sentence makes clear, and his correspondence with Hilbert even clearer, Frege never really abandoned a Kantian view of (Euclidean) geometry, whose statements he regarded as synthetic a priori. As a result, it would seem highly implausible that he would have come to hold a view of philosophy similar to Russell’s, at least insofar as Russell’s own view was primarily motivated, as I want to claim here, from his compensation objection against Kant.

5 It is also worth noting that when Russell first wrote these words, he was not yet familiar with Frege’s work.
very defective, and that the philosophical logic based upon the syllogism showed a corresponding narrowness. What is now required is to give the greatest possible development to mathematical logic, to allow to the full the importance of relations, and then to found upon this secure basis a new philosophical logic, which may hope to borrow some of the exactitude and certainty of its mathematical foundation. [my emphasis]

Of course, one has to take into account that Russell is being programmatic here (if not propagandistic) and that the period during which these words were written comprises perhaps the height of Russell’s optimism about what had been achieved with the *Principia*. Nevertheless, I do believe this passage contains the core founding idea of analytic philosophy, which is this: given a mathematical logic associated to a foundation for mathematics, one can then “found upon” it a philosophical logic that can be applied to clarify and tackle philosophical problems.

A few clarifications are now in order. Firstly, Russell speaks simply of a “mathematical logic” and does not associate it to a foundation of mathematics. This is because, from Russell’s logicistic viewpoint, a mathematical logic simply is a foundation of mathematics; since all of mathematics is reducible to logic, a mathematical logic is simply the logic to which all of mathematics is reducible. In Russell’s case, this “mathematical logic” is the logic presented in *Principia Mathematica* which is more or less what is now known as a (ramified) theory of types.

Eventually, however, the notion of a foundation of mathematics is separated from the notion of a logic, partly because of Gödel’s and Hilbert’s (and their followers’) mathematical work on formal systems, and partly because of the seemingly insurmountable philosophical obstacles one faces when attempting to describe devices such as the Axiom

---

6Russell must have approved the above view at least until 1917, the year in which [155] was included in an anthology of his writing.
of Reducibility as “laws of logic”. Thus, the notion of foundation of mathematics is no longer identified with a mathematical logic (at least for the non-logicist). Instead, we come to a picture of the foundation of mathematics in which there is a logical “core” (a “basic language” in the sense of Section 2.1 below, e.g. first-order predicate logic with equality) and a non-logical “theory” (e.g. the axioms of ZF set theory expressed using “∈”) usually expressed in terms of this basic language (though this need not be a strict constraint). In addition, one can think of this “theory” as motivated by some intuitive conception of what it is meant to describe (a “Universe of Objects” in the sense of Section 2.1, e.g. the cumulative hierarchy of sets). This logical “core” of a foundation for mathematics is what I will now mean by a “mathematical logic” (and such a mathematical logic is associated to a foundation by being its core).

In these modern terms I therefore take the relationship between foundations of mathematics and philosophy that Russell had in mind to be the following: first-order predicate logic with equality is a mathematical logic associated with set-theoretic foundations of mathematics; so a philosophical logic can be founded upon first-order logic with equality that can be applied to clarify and tackle philosophical problems. This philosophical logic, built on top of first-order predicate logic with equality (“FOL=”), constitutes what Russell called the “philosophy of logical analysis” and what we now usually refer to as Analytic Philosophy.

In other words, by “Analytic Philosophy” I will understand the discipline one of whose main methodological tenets is that FOL= with set-theoretic semantics can be applied to the investigation of philosophical questions, both in order to clarify pre-existing questions, as well as in order to state possible answers to such questions. As the discussion so far should hopefully have made clear, I also believe that this is the conception of philosophy that may plausibly be said to follow in Russell’s footsteps.
founded upon the mathematical logic of first-order logic and set theory. To illustrate, here are some common ways in which this philosophical logic is put to use:

- Using biconditionals of the form “$A$ iff $\phi$” (where “$\phi$” is a first-order formula or where the whole biconditional is a first-order formula) in order to clarify a specific thesis being argued for or to set a definition. For example, “$\forall x \forall t \ (p \text{ is good iff } (x \text{ at } t \text{ ought to desire } p) \land (x \text{ is not drunk}))$” as a potential thesis in ethics concerning the nature of the “good”.
- Using Tarskian semantics to state theories truth and reference. For example, “$\phi$ is true iff World$\models \phi$”.
- Using first-order logic as a means of stating the real underlying logical form of everyday language. For example, the seminal analysis of predication in *On Denoting*.
- Using set theory as an ontological greengrocer in order to clarify talk of properties and concepts and in order to state theses concerning ontological dependence. For example, the use of the distinction between Socrates and the singleton \{Socrates\} in recent literature on “grounding”.

Thus, what was new in the “philosophy of logical analysis” as Russell envisioned it was neither the idea of *analysis*, nor the idea of this analysis being carried out in terms of *logic*. Both ideas predate Russell by millenia: the idea of philosophy proceeding by the analysis of concepts has been present, in one form or another, since the very beginning of the Western canon; and the idea that some of these concepts can be fruitfully analyzed in terms of their logical form has been around at least since Aristotle. What is new in the Russellian conception of a philosophy of logical analysis is the particular mathematical logic with which this “logical analysis” is to be performed.
Therefore, if one could find an alternative mathematical logic associated to an alternative foundation of mathematics, then there is just as much reason (even by Russell’s lights) to build a “philosophical logic” on top of this alternative logic as there was to build the philosophical logic that we now call Analytic Philosophy on top of first-order predicate logic with equality as it was associated to set-theoretic foundations.\(^8\) But what exactly could such an alternative foundation of mathematics be?

### 1.2. A New Kind of Synthetic Geometry

At around the same time as Cantor, Frege and Russell were laying out the groundwork for the developments that would eventually lead us to ZFC\(^9\) – and before Hilbert’s famous axiomatization of Euclidean geometry in 1899 – a very different debate was still very much alive in the mathematical world, concerning the fate of geometry. This debate focused on *synthetic* vs. *analytic* conceptions of geometry; superficially, it was about whether geometrical figures (and the spatial intuitions they encode) should play an ineliminable role in geometrical exposition and demonstration, or whether all geometry was simply about the (algebraic) manipulation of certain equations in terms of which such figures could be analyzed.

Any kind of “analytic/synthetic” distinction is likely to result in a terminological quagmire, even in the context of geometry. Some clarification as to how I will be using these terms is therefore in order. To do this it is helpful to sketch (very roughly) the history of the debate. It begins with Euclidean geometry, which was *synthetic*: its basic objects (lines, points) were irreducible things into whose nature one is supposed to have

\(^8\)This is, at best, a *historia abscondita* of analytic philosophy. I am ignoring many complexities and many key figures in merely taking Russell’s pronouncements as the founding dogma of analytic philosophy. But this narrative is sufficient for my purposes here – and I do not think that the key figures it omits, e.g. Moore – contradict it in any way.

\(^9\)Although I have grouped Cantor, Frege and Russell together, it must be made clear that Cantor hardly saw himself as engaged in the same kind of project as Frege – indeed, Cantor seemed to have disdained formal systems altogether as an approach towards the foundations of set theory, explicitly distancing himself from Frege’s *Begriffsschrift*. See [35], pp. 221-224 for a relevant discussion.
some kind of privileged access. One chief problem with Euclidean geometry was its seemingly feeble claim to universality: though demonstrations were rigorous, each particular figure had to be given in one sense or another (e.g. drawn or scrawled). Whence could one conclude that a certain demonstration held for arbitrary triangles if it could only be carried out on a given triangle? And indeed, the diagrammatic reasoning of Euclid appears to go wrong even at the very first demonstration in Book I.

Cartesian geometry offered a way out of the conundrum. By liberating geometry from the particularity of figures it infused it with the universality of arithmetic by making precise in exactly what sense a given triangle was arbitrary, viz. by satisfying a certain prescribed equation. Kant, as we saw, then tried to restore to mathematics its synthetic nature without sacrificing its universality by grounding its propositions in the synthetic a priori: although any demonstration of a particular fact about a triangle required the construction of a given triangle, this construction took place a priori in pure intuition. This, according to Kant, gave any of the demonstrations carried out on such a construction universal validity, thus vindicating a synthetic viewpoint of Euclidean geometry.

In the beginning of the 19th century, with the discovery of non-Euclidean geometries, the Kantian picture comes under relentless scrutiny. Quite naturally, this has important repercussions on the debate between synthetic and analytic geometry. Decreased reliance

---

10Its basic objects were, of course, defined (“A point is that which has no parts”) but these definitions played no part in the deductions.

11See [27] for a discussion of how Euclid’s demonstration fails modern standards of rigour. Burgess’ historical account is “Whig”, as he himself confesses, in that he does not hesitate to read our modern understanding of Euclid into the ancient text. In particular, Burgess assumes that the diagrams themselves play no role other than being illustrations of the demonstration at hand, adding nothing to the logical structure of the argument. Others, most notably Manders [118], have tried to rescue Euclid by understanding the diagrams he uses not as mere illustrations, but also as a kind of notation governed by its own rules. But in any case, Burgess is certainly right that people generally perceive Euclidean geometry to fail contemporary standards of rigour.
on geometric intuition together with significant developments in algebra mean that analytical methods in geometry come to dominate over diagrammatic ones.\textsuperscript{12} Combined with the sustained philosophical attacks against Kant, towards the end of the century the Euclidean idea of a synthetic geometry based on diagrammatic reasoning falls into near-total disrepute. Indeed, the very term “synthetic” seems to altogether change meaning, even in the context of geometry, and becomes associated with the “axiomatic” method pioneered by Hilbert, wherein geometrical postulates are not taken to describe any geometrical notion in particular, but rather to \textit{implicitly} define such notions regardless of whether we wish to regard them as geometric or not.\textsuperscript{13}

A lesser-known aspect of this story is that from the ashes of this debate a new kind of synthetic geometry is reborn through the work of figures such as Riemann, Grassmann, Möbius, Klein and Poincare. This new synthetic geometry is dominated by the idea of \textit{invariant form}, i.e. of geometric spaces, shapes and figures determined only \textit{up to} some shared (algebraically expressed) features. Crucially, it involved elements of both analytic geometry in the Cartesian tradition and of synthetic geometry in the Euclidean tradition, but could not be identified with either. Its origins are perhaps best traced back to a letter Leibniz wrote to Huygens in 1679, in which he described his idea for a “Geometric Characteristic” which would “have great advantages for representing to the intellect everything that depends on the imagination precisely and naturally, but without figures.”\textsuperscript{14} Leibniz’s idea was never fully worked out, neither by him nor by any of his contemporaries. As a result, it did not attain a state of development any more advanced

\textsuperscript{12}The association of “algebra” with “analytic” methods may appear strange to the modern ear, but it is very much standard in the 19\textsuperscript{th} century. Consider, for example, the way Helmholtz [182] refers to the “analytical method of modern algebraical geometry” (p. 668).

\textsuperscript{13}The clash between the two senses of “synthetic” is very forcefully brought out in the (in)famous Frege-Hilbert correspondence, in which neither man seems capable of understanding that they are talking past each other.

\textsuperscript{14}\textit{GM. II, 20-27}, translated by Kannenberg in [53].
than the few sketchy remarks found in Leibniz’s letters and some unpublished papers.\textsuperscript{15} In the 19th century, however, this fundamental idea of Leibniz’s is revived through the work of two key figures: Riemann and Grassmann.

Riemann’s contributions to mathematics are so deep in their impact and so immense in their width that it seems scarcely possible to talk about contemporary mathematics at all without mentioning his name. The contribution that most interests us here is Riemann’s remarkable paper “On the hypotheses which lie at the foundation of geometry” (cf. [152]) which he read out to an audience of philosophers and mathematicians (including Gauss) in 1856. In this paper, Riemann establishes the definition of the modern notion of a manifold, thus being the first to set out a clear distinction between the notion of a space and the notion of a geometry. Non-Euclidean geometries, Riemann argues, emerge as a perfectly natural consequence of the fact that there are many incompatible geometries one can impose on three-dimensional space – and that whether or not the real space of our sense perception is or is not Euclidean is a matter best decided by empirical, not mathematical, investigation:

[T]he propositions of geometry cannot be derived from general notions of magnitude, but [...] the properties which distinguish space from other conceivable triply extended magnitudes [three-dimensional manifolds]

are only to be deduced from experience. ([40], p. 652)

In doing so, Riemann sets the stage for the study of space abstracted from its geometric properties and almost single-handedly gives birth to the field of topology: the study of

\textsuperscript{15}A possible earlier source for this kind of idea is Euclid himself. In Book II of his Elements, Euclid seems to give demonstrations to “obvious” geometric facts that can be seen (from a modern perspective) to encode well-known algebraic identities. Indeed, there was a popular line of interpretation that took Euclid to have used geometry as a way of doing algebra, i.e. of regarding diagrams themselves as playing the role of algebraic formulas. This is almost the converse of the Leibnizian idea but at the end of the day they amount to the same suggestion: an organic synthesis of algebra and geometry. This point of view on Euclid, although popular with mathematicians, was highly controversial and indeed led to an infamous dust-up between the historian Unguru [170] and the mathematician Andre Weil [184] (among others) played out in the pages of the journal Archive for History of Exact Sciences.
spaces without a notion of “distance” or “length”. Through Poincare’s recognition that the Betti numbers of a topological space can be calculated as the dimensions of certain algebraic structures, one obtains the modern field of algebraic topology. And in algebraic topology one studies spaces very much in the spirit of the Leibnizian characteristic: there is invariant, irreducible geometric form but this form is expressed through algebraic structures subject to analytical-logical operations, not diagrammatic reasoning. Needless to say, Riemann’s contribution, though essential, is also unconscious: it seems unlikely that he was at all motivated by the Leibnizian characteristic in his own investigations.

Grassmann, on the other hand, explicitly regarded his own Ausdehnungslehre \([53,54]\) as the realization of the Leibnizian characteristic. In his prize-winning Geometric Analysis of 1847, Grassmann believes that he has

>[formulated], at least in outline, an analysis which in general actually accomplishes what [Leibniz] regarded as the goal of [his Geometric Characteristic]. ([53], p. 318)

In Grassmann’s Ausdehnungslehre, this basic idea is then to be applied not only to geometry, but to the whole of mathematics:

Pure mathematics is [...] the science of the particular existent that has come to be by thought. The particular existent, viewed in this sense, we call a thought form or simply a form; thus pure mathematics is the theory of forms. ([53], p. 24)

It is here that we find the essential kernel of a revolutionary idea that is entirely independent of Cantorian set theory, or of the foundational developments in the 20th century that it inspired, namely the idea that all of mathematics can be encoded in terms of forms that have some kind of intrinsic spatial meaning. It is this idea that structuralist
foundations like the Univalent Foundations are finally able to do justice to, as we shall see.

But what exactly does such an idea entail? If all of mathematics is to be understood as the study of forms with intrinsic spatial content, then we want “thick” basic objects with a lot of native structure as opposed to the “foundationalist” approach of building everything up from maximally “thin” basic objects (e.g. $\emptyset$). What this idea requires is a formal system in which the basic objects can be understood as spatial entities of some kind, described at a level of abstraction that makes them amenable to logical (rather than diagramatic) manipulation. In other words: a Begriffschrift for shapes. In describing (his own exposition of) Grassmann’s work, Peano in his Geometric Calculus explains very well what this idea would entail:

The geometric calculus, in general, consists in a system of operations on geometric entities, and their consequences, analogous to those that algebra has on the numbers. [...] [It] exhibits analogies with analytic geometry; but it differs from it in that, whereas in analytic geometry the calculations are made on the numbers that determine the geometric entities, in this new science the calculations are made on the geometric entities themselves. ([147], p. ix) [my emphasis]

Thus, such a “geometric calculus” is analytic in the sense that only logical operations are to be performed on its basic entities; but it is synthetic in the sense that these basic entities possess some irreducible and non-trivial form. Furthermore, and more mysteriously, it is analytic in the sense that it relies on formulas rather than figures; but it is synthetic in that these formulas themselves have geometric content, which was also the kernel of Leibniz’s idea.\footnote{The idea of “formulas themselves having geometric content” will be explored further in Chapter 3.}
But in what way exactly such an idea could materialize into a usable formal system – and a foundational one no less – remained unclear until very recently.\footnote{Of course, at around Grassmann’s time even the very idea of a “formal system” was lacking in the precision which it now enjoys. But Peano certainly had a conception close to the modern one and he certainly believed it was applicable to Grassmann’s system, as evidenced by his [147].} For even before we can ask how all of mathematics is supposed to be encoded in terms of primitive geometric forms, until very recently it remained unclear, beyond Grassmann’s murky descriptions, exactly what could play the role of these forms. And the answer – that they should be homotopy types – emerges, as I hinted at, directly out of the branch of mathematics that begins in its mature form with Riemann, namely the field that is now known as algebraic topology.

However, all these incomplete suggestions that emerge out of Riemann and Grassmann existed a great distance away from the ideas for a foundation of mathematics as they were being developed by Frege and Russell. Indeed the idea that homotopy types could have anything to do with the foundations of mathematics would have seemed outlandish even as recently as a decade ago. Rather, the immediate impact of this strain of synthetic geometry was in spurring the mathematical development of “abstract” algebraic geometry and topology.\footnote{Of course, the discovery of non-Euclidean geometries was perhaps an even greater spur for these developments and so was Einstein’s application of many of them to his formulation of the general theory of relativity, though that came much later.}

Felix Klein’s much-heralded Erlangen program – which aimed to classify different geometries according to which symmetries the properties of its basic figures were invariant under – then emerged as perhaps the most mature expression of this new kind of synthetic/algebraic geometry, in which primacy is given to the geometrical content of algebraic descriptions. And category theory – invented by S. Mac Lane and S. Eilenberg in the first half of the 20th century – then offered a framework in which to develop Klein’s
ideas in a vastly more general vista, thus giving rise, eventually, to the monumental re-
imagining of algebraic geometry in the 20th century by Grothendieck and his school. The new structuralist foundations that this dissertation will be concerned with emerge
directly out of the category-theoretic approach to algebraic geometry that was pioneered
by Grothendieck and his school. As such, they are directly related to the renascent idea
for a synthetic geometry that originates with Grassmann and Riemann and this historical
link is crucial in understanding them.

To summarize: from Grassmann we get the idea that all of mathematics can be
thought of as the study of spatial forms; from Riemann we get the field of mathematics
that studies spatial forms of this kind. Through this nexus of ideas, and primarily via
Grothendieck and category-theoretic methods, emerges the field of algebraic geometry
that gives rise to the modern notion of a homotopy type. Yet all these seem like purely
mathematical developments. How exactly, one might now reasonably wonder, does all
this relate to philosophy?

1.3. Univalent Foundations and Philosophy

Somewhat surprisingly, mainstream analytic philosophy and category-theoretic math-
ematics have had little interaction. As such, philosophy has gone about its business with-
out much regard for the methodological developments in mathematical thinking in the
20th century that led to this new kind of synthetic geometry outlined in Section 1.2. This
attitude, however, was entirely justified by the following fact: category-theoretic,
Erlangen-style, Grothendieck-inspired mathematics was still developed and made precise

19Marquis [120] has made a compelling study of the connections between category theory, Kleinian
geometry and Grothendieck’s vision for algebraic geometry.

20Beyond the philosophy of mathematics, or issues directly related to mathematics (e.g. [32]), to my
knowledge the only project in analytic philosophy that aims to directly apply category theory to pre-
eexisting philosophical debates is the one outlined and pioneered by Halvorson in [57]. For further
elaboration on Halvorson’s project see also [58].
in terms of set theory.\textsuperscript{21} Algebraic geometers of the 20th century – Grothendieck included – still very much thought of set theory as the foundation of their mathematics and were more than happy to carry on inhabiting Cantor’s paradisal gardens.\textsuperscript{22} For example, even if what one cared about was some sort of equivalence class of spaces rather than a \textit{given} space, this equivalence class of spaces was still defined \textit{on} sets: one started with something that had too much specificity (e.g. $\mathbb{R}^4$) and systematically determined how much of this specificity was to be considered redundant. And since philosophy cares primarily about the foundations of knowledge and not about cutting-edge methods we have of \textit{obtaining} it, it was still reasonable for philosophers to think of mathematics in terms of sets and deductions in the first-order theory of sets. From this point of view it was just as reasonable for philosophers to disregard developments in category theory as it is for particle physicists to disregard developments in mechanical engineering.

Thus, the new conception of geometry outlined in Section 1.2 had seemingly little role to play in philosophy. With the arrival of the Univalent Foundations, this attitude ceases to be reasonable. For the basic objects of UF – homotopy types – play exactly the role of the invariant forms of this new kind of synthetic geometry we have described: homotopy types can be profitably understood as a vast generalization of Grassmann’s “invariant forms” in his \textit{Ausdehnungslehre} \textsuperscript{53, 54}.\textsuperscript{23} Indeed it is no stretch to see UF as a modern-day version of the \textit{Geometric Calculus} of Peano \textsuperscript{147}, which was originally intended to provide a logical foundation for Grassmann’s ideas. Crucially, UF’s intended

\textsuperscript{21}For example, even Bourbaki’s work on the foundations of mathematics still very much relied on axiomatic set theory. See Corry \textsuperscript{33} for a helpful overview.
\textsuperscript{22}Even Grothendieck, one feels almost against his will, fell back to set theory in order to give his reimagined algebraic geometry firm foundations. As Marquis \textsuperscript{120} notes: “in dealing with foundations, Grothendieck fell back on set theory.”
\textsuperscript{23}See \textsuperscript{121} for a philosophical take on homotopy types, and how they relate to the movements in 20th century mathematics alluded to above.
semantics lie entirely outside the spectrum of what we may call “set-theoretic foundations” of mathematics in the Cantorian tradition. This point is put very nicely by Manin when he compares homotopy types with Cantorian sets:

Instead of sets, clouds of discrete elements, we envisage some sorts of vague spaces, which can be very severely deformed, mapped one to another, and all the while the specific space is not important, but only the space up to deformation. Earlier, all these spaces were thought of as Cantor sets with topology [...] I am pretty strongly convinced that there is an ongoing reversal in the collective consciousness of mathematicians: the right hemispherical and homotopical picture of the world becomes the basic intuition, and if you want to get a discrete set, then you pass to the set of connected components of a space defined only up to homotopy.

([50], p. 1274)

In Section 1.1 I argued that there is nothing more to first-order logic and set theory that make them suitable to provide the philosophical logic with which analytic philosophers work today, other than the fact that they are associated with a certain conception of the foundation of mathematics, one that Russell was instrumental in developing (even if the currently-accepted version (ZFC) does not coincide with his). And what I want to say now is that the Univalent Foundations – based as they are on an entirely non-Cantorian semantics – should just as well be able to provide us with a new mathematical logic and therefore a new philosophical logic.

Now, one might object that the new foundations in the early 20th century did not merely formalize objects, but also patterns of inference that had up to that point eluded such formalization. For first-order logic was capable of formalizing not only mathematical *notions* (e.g. continuity) but also mathematical *inferences* that Aristotelian logic
could not (e.g. existential instantiation or universal generalization). And, arguably, the expansion of formalized inferences is of more importance philosophically than the formalizations of previously-vague mathematical concepts. Thus, we must now ask: can these new foundations achieve such an expansion too? Do they succeed in formalizing yet-unformalized patterns of inference in mathematics?

They do indeed. The pattern of inference that they formalize is that of transferring properties and structures of one object to any object isomorphic to it.\(^{24}\) This is a pattern of inference ubiquitous in contemporary mathematics. It is invoked under many names in (informal) mathematical practice, usually referred to as “abuses of notation”. For example, one identifies the rational numbers \(\mathbb{Q}\) with their image \(i(\mathbb{Q})\) in \(\mathbb{R}\) under the canonical injection \(i: \mathbb{Q} \hookrightarrow \mathbb{R}.\)^\(^{25}\) Strictly speaking, however, these two entities are not equal as sets, even though “clearly” anything we say of one we should also like to assert of the other. But this “obvious” fact cannot be formalized in set theory. Indeed it is invalid as an inference since as long as we regard \(\mathbb{Q}\) and \(i(\mathbb{Q})\) as distinct (though canonically isomorphic) sets there will be (first-order) formulas in set theory true of \(\mathbb{Q}\) but not of \(i(\mathbb{Q})\).

Of course the usual response to this, by set-theorists and mathematicians alike, is simply that “We know how to do this if the need arises” or some variant of the complaint that “It is intuitive – it would be mere pedantry, if not bad taste and boring, to try and formalize these inferences so that they become valid, since we already know it can be done in principle.” But one need only recall that exactly analogous statements were being made in the 19th century in the context of analysis (concerning the notions of continuity, smoothness, limits etc.) and that attempts to rigidly and formally demarcate

\(^{24}\)This is achieved through the axiom of univalence. I will have much more to say about that in Chapter 2.

\(^{25}\)What this “canonical injection” does exactly depends of course on how we have chosen to define \(\mathbb{Q}\) and \(\mathbb{R}\).
them were often dismissed as pedantry unworthy of the practicing analyst; yet it is these attempts exactly that led to set theory in its present, fully mature, form. The situation with respect to the above-described pattern of inference is thus non-trivially analogous to the situation in the 19th century. Practicing mathematicians now argue all the time with the understanding that proofs and arguments can be repeated (i.e. “transferred”) along isomorphisms, leaving it implicit, and regarding it as obvious, that those proofs and arguments and statements that cannot be transferred in this manner are not the ones we “care about” anyway. This is eerily similar to how mathematicians of the 19th century thought that it was obvious which functions were continuous and which were not – and that to make precise a criterion that could distinguish between them would be to indulge in tasteless pedantry.

But closer to our concerns here, this means that there is indeed a very-well understood pattern of inference, namely “transferring properties and structure along an isomorphism”, that is accepted and used extensively in the practice of mathematics, but which no formal system has yet managed to formalize. There were primarily two technical hurdles that had to be overcome in order to achieve this formalization. Firstly, to come up with a general notion of “isomorphism” which specializes to most of the cases of “isomorphism” in mathematical practice. Secondly, to come up with a formalism that treats identity as a structure rather than a proposition. Both these hurdles, as we shall see in Chapter 2, are overcome in the Univalent Foundations: homotopy equivalence serves as the general notion of isomorphism, and the identity types of dependent type theories are able to formalize identity as a structure, rather than merely as a proposition with a truth value. Therefore, the Univalent Foundations do not merely offer an

26For evidence of such dismissive attitudes towards rigorization consider Charles Hermite in an oft-quoted 1893 letter to Stieltjes: “I turn with terror and horror from this lamentable scourge of continuous functions with no derivatives.”
alternative conception of the basic objects of mathematics, but also manage to formalize
previously-unformalized patterns of inference that are of use in mathematical practice.

One might also object that to ignore Russell’s logicism is to throw out the baby with the
bathwater. For what mattered to Russell was not merely that a mathematical logic
could be used to produce a philosophical logic, but rather that this mathematical logic was
Logic itself, i.e. a logica magna expressing immutable laws rather than one among many
possible “core” formal systems that could be associated to a foundation of mathematics.
After all, Russell writes that the “fatal blow” to the Kantian philosophy is dealt by the
fact that “all pure mathematics, including Geometry, is nothing but formal logic”. If,
as I have suggested, following the modern point of view, we dissociate mathematical
logic from foundations (i.e. “basic language” from “theory”) then it is unclear if Russell
would still be willing to make the connection between formal logic and philosophy that I
have interpreted him as making, or that the connection between analytic philosophy and
foundations of mathematics that I have drawn would still stand.

There are a few things I have to say in response to such an objection. Firstly, one
certainly need not espouse a logicist position in the philosophy of mathematics in order to
call oneself an analytic philosopher. The idea that formal logic may have an application to
philosophy can certainly be dissociated from the idea that all of mathematics is reducible
to formal logic. Secondly, Russell himself seemed open to rethinking the whole formal
system he was working with, as evidenced by the many revisions he allowed the system of
the Principia Mathematica to undertake between the first and second edition (following
criticisms of Wittgenstein, Chwistek and Ramsey, for example), his eventual abandon-
ment of the Axiom of Reducibility, as well as his willingness to declare even as late as
1920 that “much of [the theory of types] is still inchoate, confused and obscure.”27 As

27[156], p. 135.
such, I do not think it at all implausible that Russell would recognize some “core” mathematical logic that remains more or less constant throughout such reconsiderations of the larger system. Finally, independently of what Russell may or may not have thought, I do still find the analogy I am drawing plausible even without some *logica magna*-style view of formal logic. What happens in the foundations of mathematics is that a certain core mathematical logic is employed in order to describe all of mathematics in as precise and coherent manner as possible. Exactly analogously, what happens in analytic philosophy is that certain core patterns of reasoning (considered *rigorous*) are employed in order to describe all of the world in as precise and coherent manner as possible. It therefore seems to me entirely plausible to assert a connection between the core patterns of reasoning in philosophy and the core mathematical logic of a given foundation of mathematics.

To summarize: I have argued that the move from set-theoretic foundations to the Univalent Foundations is of comparable importance to the move from the early attempts at rigour in the 19th century to the formal systems that are now called foundations of mathematics (with ZFC as the paradigm) – for not only does UF provide a new conception of the basic objects of mathematics but it also formalizes previously unformalized patterns of inference. Furthermore, I have explained that the connection between mathematical logic and philosophical methodology that founded analytic philosophy does not require the espousal of some strong background logicism, comparable to Russell’s. All it requires is a foundation of mathematics and a mathematical logic associated to it. The way forward is then very simple: I need to describe and clarify a new foundation in such a way as to obtain a new mathematical logic that is associated with it and then “found upon” this mathematical logic a new philosophical logic. This is exactly what this work intends to do: the description and clarification of the Univalent Foundations is the business of Chapters 2-4 and it leads up to the new mathematical logic developed in Chapter 5.
1.4. A Challenge for Recent Metaphysics

But before going forward, let me first go backward and examine what such developments in the foundations of mathematics may have to say for the present-day analytic philosopher. The reliance of contemporary analytic philosophy on set theory is near-total. This is especially true of metaphysics and the philosophy of language, but no “sub-discipline” has really managed to escape, even those not traditionally associated with formal methods. As some kind of empirical evidence for the ubiquity of set-theoretic parlance (and intuition), consider these quotations from (distinct) papers of David Lewis collected in [97]:

I understand you now: co-perforation is supposed to be an *equivalence* relation among hole-linings [...] [my emphasis] ([97], p. 7)

The standards of validity for modal reasoning have long been unclear; they become clear only when we provide a semantic analysis of modal logic by reference to possible worlds and to possible things therein.\(^{28}\) ([97], p. 10)

I wish to regard enduring things such as persons and bodies as aggregates – *sets*, mereological sums, or something similar – of momentary stages. ([97], p. 47)

The definitive causal role of an experience is expressible by a *finite set* of conditions that specify its typical causes and its typical effects under various circumstances. ([97], p. 102)

\(^{28}\) There is a footnote at this point in Lewis' text which refers the reader to Kripke's classic paper [82] on the semantics of modal logic, which of course does not give an intuitive semantics but a formal semantics in set theory. The distinction between pre-formal semantics and formal semantics is discussed in more detail in Section 3.3.
A grammar, like a language, is a set-theoretical entity which can be discussed in complete abstraction from human affairs. ([97], p. 176)

I consider Lewis to be the quintessential example of an analytic philosopher in the tradition brought about by Russell as described in Section 1.1, as well as a paradigm case of a systematic philosopher who believed that every philosophical question could be resolved if he were granted certain ideas and formal tools. Lewis is therefore put forward here as the model target for the criticism of analytic philosophy that I want to present in this chapter, viz. as a systematizer whose philosophy took the resources of set theory for granted.

I take every one of the above passages to rely on set-theoretic intuitions and concepts that do not carry over to the basic objects of the new foundations that I will be considering. This does not, of course, prove them wrong – nor does it prove Lewis misguided. The question it raises, rather, is what exactly is it that gives analytic philosophers such confidence to employ set theory and first-order logic in their formal philosophizing? If some intuitions associated with sets do not survive the transition to the new foundations, does that not constitute a reason to question philosophical doctrines that have been built upon them?

Such questions are especially pressing for metaphysics, to which I will now turn my attention. The dilemma I want to pose for analytic metaphysics is this: if the philosophical logic that analytic metaphysics depends on is only one out of many equally plausible alternatives, then what licenses our blind faith in set theory and first-order logic? And if this faith is not blind, then what is it exactly that justifies it? In other words, it seems to me imperative for the analytic metaphysician, given these new developments in the foundations of mathematics and the connection I have drawn between foundations and analytic philosophy, to provide one of the following:
(1) *Either* a metaphysical justification for the primacy of set theory and first-order logic as a formal framework for metaphysics

(2) *Or* some demonstration that the doctrines of analytic metaphysicians do not essentially rely essentially on set theory and first-order logic.

I believe the first challenge is impossible to answer for the reason that it requires metaphysics *prior* to set theory; and a metaphysics that can justify set theory as a framework for metaphysics would itself already provide the required framework for metaphysics. I will now say a few more words about this here.

The obvious way out for the metaphysician is to claim that set theory does indeed enjoy some kind of privileged ontological status as compared to all other possible proposals for a foundation of mathematics. If this is established, then certainly new foundations have nothing to add to current debates on metaphysics, nor will they undermine any of the fundamental examples that rely on set theory. One simple way of achieving this is to espouse some kind of set-theoretic Platonism, so that one can take facts about the cumulative hierarchy to be facts of nature in a very strong sense. Such a move would of course license the reliance of metaphysicians on set theory since it would be no different to their reliance on modern theories of physics. But needless to say, set-theoretic Platonism would then provide a whole new tar pit to navigate.29

On the other hand, another plausible way to proceed, independent of set-theoretic Platonism, is to piggyback on some claim about the primacy of first-order logic with equality (FOLₑ) as a formalization of deductive reasoning, i.e. to take set theory’s primacy to derive from the fact that it is the intended and most natural semantics for FOLₑ.

29 In addition, I suspect that less controversial and ontologically taxing “thin” conceptions of set-theoretic Platonism (e.g. Maddy’s “Thin Realism” in its most recent form elaborated in [107]) are not themselves “thick” enough to do the work that, say, a grounding theorist would require of them; by Maddy’s own account, for instance, her view seeks a “post-metaphysical objectivity” and grounding seems very much to be concerned with metaphysical objectivity in the strongest possible sense. In other words, if one tries to dodge the bullet by attempting to rely on a non-metaphysical conception of set-theoretic Platonism, one risks altogether bypassing the real ambitions of metaphysicians.
The status of \( \text{FOL}_m \) as a logic is of course an issue of great philosophical complexity and interest and one we can hardly even broach at this point. But even if we grant that \( \text{FOL}_m \) enjoys some privileged status as a “fundamental calculus of thought” I still don’t think the metaphysician can get away with such a strategy, for three reasons.

Firstly, it does not follow from the fact that \( \text{FOL}_m \) is a fundamental calculus of thought that it is the fundamental such calculus. Indeed, it is difficult to make sense of what such a thing would be since, qua calculi, there are very many ways in which to interpret formal systems one into the other. A famous and counterintuitive such example is Gödel’s \( \neg\neg \)-translation of classical \( \text{FOL}_m \) into intuitionistic \( \text{FOL}_m \), which indicates that if we regard classical and intuitionistic logic merely as formal systems then the question of fundamentality seems altogether lacking in sense, since each is capable of simulating the laws of the other. Therefore, even if we grant that \( \text{FOL}_m \) is a fundamental calculus of thought, it cannot meaningfully be regarded as the unique such; and insofar as there can be other such fundamental calculi (such as the one I develop in Chapter 5 backed by the intuitive explanation of Chapter 3) the metaphysician is still making an arbitrary choice in picking \( \text{FOL}_m \) alone, ignoring other calculi which encode possibly conflicting intuitions.\(^30\)

Secondly, there is not necessarily any good reason that the fundamentality of a formal calculus should carry over to its “intended and most natural” semantics. For example, possible worlds might indeed be the intended and most natural (set-theoretic) semantics for modal logic, but it seems a stretch to go from that to the assertion that (set-theoretic) possible worlds are fundamental to our modal reasoning.\(^31\) Or, to use a less philosophical,

\(^{30}\)What I have said here clearly depends on a specific view of logic, one that does not take it to be a “logica magna” (to borrow a phrase of Moore in [140]). Just as with the set-theoretic Platonist, I say that anyone holding such a view of logic is already in possession of a fundamental framework for metaphysics, and need not bother enter our little debate at all.

\(^{31}\)Of course, many, if not most, analytic philosophers have done so. See for example Lewis’ quotation above ([97], p. 10). It is interesting to note, however, that Kripke himself very clearly distances himself
but more technical and striking, example: it is true that certain non-monotonic logics (linear logic) are sound and complete with respect to semantics in string diagrams (knots), but it hardly seems plausible to claim that that knots are somehow fundamental to understanding non-monotonic reasoning.

Thirdly, it is doubtful that the issue of whether $\text{FOL}_\leq$ precedes set theory is as clear-cut as our modern axiomatic conception of theories has led us to believe. In the early days of set theory it would certainly be reasonable to regard $\text{FOL}_\leq$ as the derived “internal logic” of certain standard set-theoretic operations. For example, consider the point of view expressed by Weyl:

[First-order] logic was abstracted from the mathematics of finite sets and their subsets [...] Forgetful of this limited origin, one afterwards mistook that logic for something above and prior to all mathematics, and finally applied it, without justification, to the mathematics of infinite sets. This is the Fall and original sin of set-theory. ([185], p. 10)

The view of logic as that in terms of which mathematical statements are made, rather than just as one more sub-discipline of mathematics, has come to dominate the analytic tradition. But the Boolean-Piercean tradition of algebraic logic is certainly one in which the whole of $\text{FOL}_\leq$ can be conceived, and in such a way that there is really no clear sense in which it precedes set theory.

Now, with respect to the second challenge, the metaphysician may claim that sets and first-order logic really do not enter into metaphysics at all. Metaphysics deals with concepts such as properties, entities, haecceities, qualities, substances etc. First-order logic and set theory is merely the grammar this particular tradition (analytic philosophy) has adopted in order to talk about these things. But these things have nothing to do from such a conclusion in Naming and Necessity: “Possible worlds’ are stipulated, not discovered by powerful telescopes.” ([83], p. 44)
with set theory – for example, properties are not themselves sets and entities are not elements of sets. Sets and their elements are merely a way we have of speaking of such things. Metaphysics is no more constrained by first-order logic and set theory than it is constrained by whether it takes place in English or German.

If this is indeed true, and if every metaphysician believes herself to be able to unformalize their arguments into non-FOL and non-set-theoretic vernacular, then my entire criticism here has missed its point. But I think this is highly implausible – and indeed for some views (e.g. views in which the world is, say, comprised of “sets of spacetime points”) I think also it is plainly false. Sets themselves have become part of the ontological arsenal of the metaphysician, not placeholders for traditional notions to which metaphysicians have always helped themselves with. (When a metaphysician says “A property is a certain kind of set” she does not take herself to be stating an analytic truth.) As for regarding the metaphysician as a kind of “set-theoretic nominalist”? Well, insofar as any kind of metaphysics can proceed nominalistically then this is a sensible position – but it is also one that cannot be accompanied by any serious attachment to set theory and first-order logic.

Thus, both horns of the dilemma pose serious difficulties. But that would be a non-issue unless there were actual widespread intuitions in analytic philosophy that would be challenged by the new foundations. I believe there are; to illustrate, I will examine a relevant issue related to the relation of “grounding” on which a lot of recent discussion in analytic metaphysics has focused (see e.g. [44, 153] for expository articles and assessments of the recent literature). Indeed, the project of coming up with “grounding explanations” has been proposed as a kind of overarching philosophical methodology,
meant to replace the method of “conceptual analysis” in the style of the so-called “Canberra School.”

A fundamental example in the literature on grounding is the example of a(n) (ur)element grounding the existence of the singleton set containing that (ur)element. This example was introduced by Kit Fine [43, 44] and nothing essential is lost if we consider as our “(ur)element” simply the empty set $\emptyset$. The example now says that the existence of $\{\emptyset\}$ is grounded on the existence of $\emptyset$ (but not the other way around). In other words, we seem to be able to point to $\emptyset$ in order to explain the existence of $\{\emptyset\}$ but it seems fruitless to point to the latter in order to explain the existence of the former, even though the latter exists if and only if the former does.

Essential to this example, and also employed in support of many of the arguments regarding the nature of the grounding relation, is the idea that Socrates is somehow “prior” to $\{\text{Socrates}\}$. This makes perfect sense set-theoretically, since the latter is obtained from the former via the power-set operation. However, this makes little sense in the new foundations of mathematics. Indeed, as I will explain in detail in Chapter 4, in the Univalent Foundations of mathematics the singleton is in a very precise sense prior to the “empty” collection. At any rate, in a philosophical logic based on these new foundations, it cannot sensibly be said that the empty “set” grounds the singleton “set”.

This is not to say, of course, that the new foundations somehow demonstrate that we got the grounding relation wrong, at least for the example of a (ur)element and its singleton. As I said in the beginning, the invocation of new formalism is not made with the purpose of substantiating new philosophical theses. I am not in the business of trying to use the new foundations in order to come up with a better definition of the grounding relation. The conclusion I wish to draw from this is that insofar as the justification for

\[32\text{For an exposition of exactly what the latter method entails, and exactly what defines the Canberra School, see [72].}\]
using set theory in metaphysics is that set theory has proved itself as a foundation of mathematics, then this justification should apply equally well to the new foundations; and in these new foundations some of the key motivating examples for grounding no longer make sense.

It may very well be the case, of course, that the grounding relation can be understood completely independent of set theory – in that sense, the intuitive notion of ground can no more be undermined by a rejection of set theory than can our faith in the proof of the infinity of primes be shaken by the discovery of an inconsistency in ZFC. Nevertheless, the reliance on fundamental examples drawn from set theory should certainly force metaphysicians to measure them up against alternative formalisms. And when they fail to measure up, as with the case of Socrates and the singleton Socrates, this ought to force a re-examination of the fundamental intuitions that these examples were meant to underwrite.

1.5. Beyond Analytic Philosophy

Everything that I have said in this Chapter has been conditional upon the existence of new foundations. The rest of (the non-Appendix part of) this dissertation will largely aim to fill in the gaps to make possible such arguments as have been put forward here. Most importantly, it will define the mathematical logic of these new foundations in such a way that a philosophical logic, pace Russell, can be built upon it. All in all, my main motivation in this work is to suggest a way of going beyond analytic philosophy (insofar as the latter is understood as tied to first-order logic and set theory).

Other such ways for rejecting or overhauling analytic philosophy have of course been suggested. In my mind, most of them belong to one of the following three categories: defeatists, naturalists and radicals. It might be illuminating to briefly examine them, as my qualms about each of them inform the key features of my own approach.
The *defeatist* stance declares that analytic philosophy can have no success in its endeavours – but that no alternative is forthcoming either. In other words, it views philosophy as experiencing its death throes, as ready to be consigned to the history books together with disciplines like astrology or alchemy – and that analytic philosophy is as good as it can ever get. As a paradigm of the defeatist attitude one may take the pessimistic meta-induction by van Inwagen, which declares, roughly, that there can be no successful arguments in philosophy because, roughly, none have yet succeeded.\(^{33}\) McGrath and Kelly \([78]\) object to this view, offering several purported successful philosophical arguments. What is interesting to me about this exchange is that both sides seem to freely conflate philosophical arguments with *proofs*, that is to say with formal deductions. This is done casually, unthinkingly, as for example here:

> For imagine that someone devises a philosophical argument that provides a relatively short or simple proof of a substantive conclusion [...] \([78\text{, p. 17}]\)

It is this conflation between philosophical argument and formal proof, I believe, that underlies both van Inwagen’s pessimism as well as Kelly and McGrath’s optimism: on the one hand, the pessimism is born of the recognition that a definitive proof of a substantive philosophical thesis cannot possibly live up to the standard of a mathematical proof; on the other hand, the optimism is born of the fact that there are trivial mathematical proofs up to whose standards a philosophical argument can certainly be expected to live up to. But why should a philosophical argument bear any similarity whatsoever to a formal deduction? Philosophical arguments as formal deductions is an ideal, of course, that analytic philosophy certainly aspires to, but one rather extreme in the demand it makes of the philosopher and altogether insensitive to the creative process of philosophy.

\(^{33}\)This view is to be found in van Inwagen’s third Gifford lecture as recorded in \([174]\).
So I think the defeatist attitude, in even expressing itself, cedes too much ground to analytic “dogma” by maintaining the idea that a philosophical argument is to be judged against the standards of a formal mathematical proof. If one rejects this piece of the “dogma” then a pessimism about the whole of philosophy based on the lack of success of its arguments no longer seems justified. And I see no reason not to reject it.

The naturalist stance believes there is hope for philosophy, but only if it looks to the empirical sciences for renewal – or, more strongly, if it adopts an empirical methodology of its very own. As an illustration of the naturalist stance, in recent work Johnston and Leslie [72] argue against a prevalent view of concepts in analytic philosophy, that in which these concepts, roughly, are understood in terms of necessary and sufficient conditions. Their observation is that such a view of concepts seems to be at odds with recent literature in empirical psychology. They censure analytic philosophers for failing to pay enough attention to certain empirical sciences; for Johnston and Leslie, to neuroscience and empirical psychology. I certainly sympathize with the intention of this kind of critique, i.e. to look for an alternative source of renewal. But I disagree on its source, namely the empirical sciences. For it certainly remains a distinctly philosophical issue, to what extent and in what way the findings of empirical science connect with the philosophical notions they supposedly help clarify (e.g. consciousness, concept-formation etc.) and insofar as this particular philosophical issue remains unresolved – or, even better, insofar as it can be shown that in broaching it one necessarily employs the very notions that are being “clarified” by the empirical sciences – then there is no ground upon which to support the analogy, and all such efforts will beg the question. In short, the applicability of empirical science to philosophy is itself a philosophical thesis, not an empirical one.

The radical stance wishes to expose as vacuous analytic philosophy’s “grand narrative” of systematic progress towards absolute truth – and thereby to reject “grand
narratives” in philosophy altogether. The rejection of a radical has more to do with the *form* of the discipline, rather than its *content*: she rejects analytic philosophy’s “scientistic” conviction that philosophy can be carried out within a grand narrative which one is meant to believe leads inexorably to absolute ahistorical truths, to *certainty*. Among philosophers who received their training within the analytic tradition, the radical stance is still perhaps best expressed by Richard Rorty. Like Rorty, the radical encourages a recognition of philosophy’s contingency within a broader cultural framework and usually, as in the case of Rorty, argues for the erasure of lines between philosophy and science and the adoption of an *ironic* stance towards one’s own (inescapable) historical situation. However sensible, meaningful, inspiring or abhorrent such a position is, I think it succumbs to a fundamental contradiction, which is that the rejection of grand narratives is itself stated in the form of a grand narrative. One may sweeten and qualify the radical stance all one likes, but it doesn’t take away from the fact that it is impossible to state it (at least within an academic context, inside the main text of a dissertation rather than tucked away in an Appendix, for example) without stating it in the *form* (a grand narrative) which it purportedly deplores. Of course, this is not to say that Rorty-like radicals are not *conscious* (and even *embrace*) this contradiction – they are not delusional, they know very well where they stand. But being conscious and explicit about inconsistent methodology can hardly quell suspicions concerning its results.

Now, let me clarify how my criticism of the three alternative approaches informs my own position. The failure of the defeatist stance reveals analytic philosophy’s much too dogmatic adherence to a methodology of problems-and-proofs, the desperate and clownish clinging on to the coat-tails of the ideal of deductive reasoning it fetishizes, and makes ever clearer the need to disentangle ourselves from the delusion of a linear progression of ideas, of an interminably gradual and slothful chipping away at a mosaic of microscopic
puzzles, each of which, we hope, will add up to the ultimate and definitive iconography of Truth and Knowledge and Science, whereas all it ever leads to is bad eyesight and a hunched back. We need to explore new forms and styles of argument and writing, beyond those that aspire to the ideal of a formal deduction.

The shortcomings of the naturalist stance reveals the need to draw from some external, non-philosophical source for inspiration and renewal, and I say we look at the foundations of mathematics, not empirical science. For in the case of mathematics, I believe there is indeed an independent and self-standing justification that allows us to import its results, concepts, and ways of thinking into philosophy. What exactly the dividing line is, between empirical science and mathematics, and exactly in virtue of what one ought to draw philosophical inspiration from the latter and not the former, I will investigate no further. My view comes closest perhaps to a persistent theme in Badiou’s work, sloganized crudely as “mathematics is ontology”, which roughly amounts to the idea that mathematical thinking is the fundamental datum (or “condition”, to use Badiou’s terminology) of metaphysics. This attitude is perhaps best-captured in [12] in which he states his guiding maxim as follows:

Philosophy must enter into logic via mathematics, and not into mathematics via logic. ([12], p. 24)

In other words, whatever deserves the name of “logic” at any given point in time emerges through the mathematics of the time, and not prior (or in spite of) it.

And the radical stance reveals that Russell’s most enduring bequest to philosophy was perhaps not his proposed method for solving long-standing philosophical problems, but the very idea that there is such a thing as a philosophical problem amenable to a definitive solution through the use of formal methods. In other words, the very idea that philosophy
can be set on the “secure path of a science”\textsuperscript{34} by imitating how mathematics can be secured by giving it foundations. The great hope of analytic philosophy, as conceived by Russell, was that formalism (in particular, formal logic) ought to constrain philosophical and metaphysical inquiry – and like all great hopes it advanced our understanding and led to fruitful, enlightening disappointments. But we must now come to recognize this great hope as deeply rooted in the soil of the scientific paradigm on which it was planted, and see philosophy not as a seed for whose growth there is one best of all possible climates, but rather as a discipline able to transcend any framework which claims that it can progress, or that there is one best climate under which it can flourish, indeed as a discipline for which there is no good progressive sense in which it can “flourish” at all. But, of course, it is not as easy as that: a distaste for grand narratives cannot be satiated by framing their rejection within an all-encompassing grander narrative. If one is to maintain some semblance of consistency, one cannot argue for the end of a grand narrative in terms of alternative research projects, governed by alternative grand-er narratives – one cannot reject the idea of a single best possible climate for flourishing by introducing the idea of a super-climate leading to super-flourishing.

In summary, the three key features that inform my suggestion for going beyond analytic philosophy are the following: contrary to the defeatist, I say we get rid of the idea that philosophical arguments should be modelled after formal deductions; contrary to the naturalist, I say we look towards mathematics for a source of methodological renewal; and contrary to the radical, I say we exhibit a new non-analytic style rather than frame it within a grander narrative of rejection. But how can all this be “argued” for? After all, I am still analyzing: this Chapter has largely consisted of a linear, leisurely argumentative

\textsuperscript{34}Ironically, a famous phrase of Kant’s.
style, with scholarly footnotes, three-fold distinctions and categorizations, careful refer-
ences, the occasional enraged digression and insult etc. In that sense, it is self-defeating,
as is the whole of this Dissertation, belonging as it does to the very same guild of pebble-
chippers that it aims to break free from. Even more laughably, it declares a distaste for
“grand narratives” in philosophy – by outlining its own grand narrative. We arrive at a
conundrum that, as I said, can only be resolved by exhibiting new form rather than new
content. To this task, reflecting the global structure of the entire dissertation but not
any of the specific points and constructions made in it, the Appendix to this Chapter is
devoted.
Appendix in the Shape of a Small Spiral

433. Let’s start now with a historia abscondita of philosophy, in the form of a cycle:

\[
\text{FORMALIZATION} \rightarrow \text{DIALECTIC} \rightarrow \text{REGIMENTATION}
\]

*Formalization* is the stage at which new terminology and notation is introduced in order to clear up contradictions and unclarities and to capture new patterns of inferences. *Regimentation* is the stage at which formalization is “brought back down to earth” thus making the new terminology and notation usable in practice. *Dialectic* is the stage at which this regimented new vocabulary once again lapses into unrigour, contradiction and unclarity, almost always through an over-reliance on a certain stock of “primitive notions”, leading once again to the need for *formalization*.

The cycle therefore is rather to be thought of as a spiral, moving ever outwards, going through each stage again and again. There are transversal movements (lines through the spiral, representing canonical or “normal” philosophy within a particular school) and transcendental movements (lines tangent to one of its curves, representing attempts at a transcendental philosophy that aims to altogether escape the particular school of philosophy that gave rise to it and set definitive ahistorical limits to inquiry.)
The transcendental movements also become transversal movements in the next cycle. The whole pattern repeats itself again and again. As the spiral grows, ever more lines converge to the center and ever more lines intersect each other:

![Spiral diagram](image)

This is a “geometric demonstration” by “diagrammatic reasoning” of the fact that the more dialectic we produce, the more meaningful things we think we can say, and the easier it is to trace an idea back to the beginning of philosophy.

353. This historia abscondita of philosophy is meant to reflect the true history of mathematics. In particular, how mathematics moves from creative semi-rigour, to formalization, to regimentation of that formalization and then once again to creative semi-rigour. The main hidden feature that I want to focus on here is what I have called “regimentation” (perhaps a better term would have been “refinement”, but it suggests “improvement” and that is not really what is going on here). “Regimentation” is meant to describe the process of making a formal language usable in normal discourse.

For example, the calculus of Leibniz and Newton can be seen as the formalization of all the talk of “infinitessimal quantities” in the natural philosophy and mathematics that preceded them. But this formalization came bundled with its own conceptual excesses, mainly due to its reliance on geometrical intuition: talk of fluxions and curves continuously approaching other curves, etc. So then in the 19th century all this was regimented into the foundations of modern analysis by Cauchy, Weierstrass and others.
by *arithmetizing* the geometric notions in the calculus, using $\epsilon$’s and $\delta$’s to express the previously wholly geometric notions of continuity. But this regimentation - aiming at the elimination of “infinites” - then began producing dialectic exactly because it lacked a firm grasp on how to handle *infinite collections* of things - and mathematics once again lapsed into unrigour, e.g. with the infamous false proof of Cauchy, later to be corrected by Weierstrass. Set theory then re-placed analysis on a firm foundation by establishing a workable notion of an “infinite collection” – and we were once again back to the stage of formalization. Thus we had the following spiral:

There are many interlocking smaller spirals emanating out of all these movements - cycles of formalization-regimentation-dialectic taking place within each movement of the bigger cycle, e.g. a cycle on surfaces (Riemannian geometry) and a cycle on logic (Boole and Pierce followed by Frege and Russell). But what is important for the particular *historia abscondita* that I am interested in is the spiral of Algebra, and how it spun off the grand movement that began with the Calculus. It can be described by the following spiral:
The dialectic of structured sets refers to the explosion of creativity in algebra in the 20th century, culminating in the proof of Fermat’s Last Theorem, and with the modern edifice of algebraic geometry, fashioned after the vision of Grothendieck. And category theory as a formal framework emerges as an attempt to formalize the reasoning that this explosion exemplifies.

281. We now find ourselves in the beginning of yet another spiral:

\[\text{Set Theoretic Foundations} \rightarrow \text{???} \rightarrow \text{Univalent Foundations}\]

although exactly where it will lead is so far not clear. From this perspective, Univalent Foundations are understood as a \textit{regimentation} of set-theoretic foundations. The Univalent Foundations will stand to set theory in the same relation that Weierstrass-style \(\epsilon-\delta\) analysis stood to the Leibniz-Newton calculus: UF aims to be a foundation of mathematics that is faithful to the current practice of mathematics, rather than to some abstract ideal of rigour that remains forever in the background.

What we are interested in then is how this movement in the foundations of mathematics should reflect on philosophy, and in particular on analytic philosophy. There is a connection between analytic philosophy and the foundations of mathematics that is often overlooked: if we see Frege (as read, improved and advertised by Russell) as the originator of analytic philosophy then we cannot remain blind to the fact that Frege was primarily concerned with the foundations of mathematics (specifically arithmetic), and \textit{not} with
philosophy. His views on sense and reference, for instance, grew out of his foundational concerns, not the other way around.

We must now use this true history of mathematics and the foundations of mathematics to try and understand what has happened in philosophy. We do this by projecting the true history of mathematics as the *historia abscondita* of philosophy. To that end, assume that movements in philosophy mirror the cycles of formalization-regimentation-dialectic in mathematics:

![Diagram of the history of philosophy](image)

Analytic philosophy can be understood as beginning with Russell reading Frege’s work on the foundations of mathematics, and then extracting from it and from his own foundations a new methodological conception of philosophy as composed of *logical problems*, *logical theories* and the capacity of these logical theories to deal with these logical problems. Implicit in this extraction is the idea that if (a) logic is capable of being the foundation of mathematics, then surely it can assume an analogous role in philosophy.

Ironically enough, philosophy did not end up adopting the particular logical calculus that Russell himself used as a foundation for mathematics (the so-called (ramified) *theory of types*). (Even more ironically, it is a type theory close in spirit to Russell’s that will cure us of Russell’s disease.) Rather, philosophy adopted what we now call *first-order logic* and in particular first-order logic *with set-theoretic semantics*. 
113. Set theory and first-order logic provide the correct framework through which to answer and settle questions. In particular, given any two propositions it becomes possible to settle the question of whether one is the logical consequence. In this syntactic task, set theory and first-order logic still reign supreme.

But set theory and first-order logic do not provide the correct framework in which to pose questions or define concepts. It is a matter altogether foreign to set theory and first-order logic whether there is a particular point to ask if \( p \) follows from \( q \).

Mathematics has known this for a long time, but philosophy, to its detriment, remains under the spell of this illusion. The disease afflicting analytic philosophy originates from a single syntactical virus, smuggled into its method as a so-called “syntactic rule” of first-order logic:

(VIRUS) Given any well-formed terms \( t \) and \( s \), \( t = s \) is a well-formed formula.

Global, unrestricted equality, and the questions it allows us to pose: this is the original sin of analytic philosophy. The virus made the leap from mathematics to philosophy—but proved far deadlier in philosophy than in mathematics. The latter became immune to it long ago. Philosophy must now also overcome it. As Weyl says, in trying to cure the paradoxes of set theory, Russell “imperiled the very life of the patient.”

It is this virus that is at the root of all the formalistic excesses of analytic philosophy, e.g. the idea of propositions as equivalence classes of possible worlds, of possible worlds as sets of facts, of theories as tuples of data, of defining a concept by giving necessary and sufficient conditions, of an agent’s actions as reducible to a set of utility functions, of thinking it sensible to ask if there is such a thing as asymmetric personal identity, if the world is made up of sets of spacetime points together with all their subsets, of mereology as set theory etc. And together with particular positions
and views, there are diseased forms of argument. In particular, the idea of giving a semantics to a syntactic construct in order to explain it, or the notion of a decisive refutation by means of a formal result (this has been Russell's most enduring bequest to the school of philosophy he spawned).

We must now call into question analytic philosophy's faith in the language of first-order logic and set theory. In historical terms: as mathematics transitions away from first-order logic and set theory as its preferred foundation, so too must analytic philosophy. To do so, the recognition and acceptance of two facts seems fundamental:

1. There are foundations of mathematics not based on first-order logic and set-theory.

2. Set-theoretic foundations are no longer capable of faithfully formalizing the post-Grothendieck style of doing mathematics.

In order to cure the virus we must return to its source: mathematical practice. We must ask: what kind of formal system formalizes actual mathematics, as it is being practiced by mathematicians?

One thing is certain: this system is not set theory. A vanishing percentage of mathematicians even know what ZFC means, or what its axioms are. (Perhaps fewer mathematicians than analytic philosophers - is that not in itself alarming?) But even if they did, that still wouldn't make a difference. For even those mathematicians that are familiar with ZFC and the project of set-theoretic foundations, do not think set-theoretically when they are actually doing mathematics, nor do they write down their proofs as deductions in a formal calculus.

To think that set theory formalizes the practice of mathematics is analogous to thinking that linguistics formalizes literature, or that optics formalizes painting. Set theory
and first-order logic do not tackle mathematical problems – they formalize them. Why should we expect them to tackle philosophical ones?

But what does all this have to do with philosophy? Because we are concerned here with the secret future history of philosophy, as suggested by the present movement in the foundations of mathematics. If we copy that movement into the history of philosophy we will get:

![Diagram](https://via.placeholder.com/150)

Analytic Philosophy → Univalent Philosophy

We are looking to find here the kind of philosophy that will stand to analytic philosophy in the same relation that Univalent Foundations (or, more generally, any structuralist foundation of mathematics) stand to set theory. I will call such a philosophy - partly to throw light on a certain period, partly to kill an hour or two - univalent philosophy.

97. The current style of mathematics requires a foundation that is faithful to it. In particular, it requires a foundation that is faithful to practice in the following sense: A distinction that is irrelevant in practice cannot be expressed in the formal language on which this foundation is based. Such a foundation we may call a structuralist foundation.

To do this one has to cut all excess fat from the formalized languages that are used in foundations, while at the same time maintaining an acceptable standard of rigour. Set theory (specifically ZFC) – the gold standard of rigour in mathematics – has failed
miserably in the former task. It has overwhelmed mathematical (and even philosophical) language with excess fat (or sand): with well-formed statements that would sound nonsensical to any practicing mathematician.

Set theory became incredibly sturdy by allowing itself an unprecedented linguistic gluttony. It thus became the fat cow that cannot be tipped over, at the cost of not being able to move at all. But we must now put the cow into a strict diet. And this not so as to tip it over, but in order to move it around, so that milk may be drawn from it by other, more advanced, methods. In particular, so it may finally be guided to the barn of Computer Science, which is to be Logic’s home in the coming century.

“There is too much sand in set theory!” Yes, indeed. And this has turned philosophy into a desert. To argue in the set-theoretic style (i.e. to do analytic philosophy) is similar to wading through sand dunes: each step requires immense exertion and it is unclear whether one has moved at all after all is said and done. But it always feels as though one is moving, even if one is not, even if one is actually sinking into quicksand.

29. Cows and quicksands aside, what we need now is a type theory. Any number of so-called “type theories” will actually have the kind of formal properties that are sufficient to make the formal and philosophical points made here. But in order for the whole vista to make any sense, this type theory must also be one that encodes a mathematical foundation faithful to mathematical practice, i.e. a structuralist foundation. And this in order to stay consistent with the particular historia abscondita which I have presented, i.e. one in which the regimenation of a particular foundation (set theory) must correspond to the regimenation of a particular philosophy (analytic philosophy).

Rather ironically, it was Russell himself that first proposed a type-theoretic system as a foundation for mathematics, albeit it was type-theoretic in a very different sense. It
involved types as a means to stratify collections of things in order to avoid the famous set-theoretic paradoxes. We now have very different reasons to adopt type theory: not in order to avoid paradox, but in order to limit identity statements, i.e. to restrict the use of equality.

And this very same Russell is also “patient zero” of the analytic virus. Having contracted it in his project to found mathematics, he transmitted it into philosophy, most famously through “On Denoting”. There, though implicit, we find that his theory presupposes a notion of (untyped) completely undetermined variables $x, y, z, ...$

I take the notion of the variable as fundamental; I use $'C(x)'$ to mean a proposition in which $x$ is a constituent, where $x$, the variable, is essentially and wholly undetermined.

and of the unrestricted formation of meaningful identity statements $x = x, x = y, ...$

The meaning of such propositions cannot be stated without the notion of identity, although they are not simply statements that Scott is identical with another term, the author of Waverley, or that thou are identical with another term, the man.

and therefore we get, from this, the syntactic rule that any sentence of the form $x = y$ is well-formed. And in this secret history, we must view this as the root of all problems in analytic philosophy. (Better: all problems specific to analytic philosophy.)

53. The formal system on which Univalent Foundations is based does not have these features: it lacks both undetermined variables and unrestricted identity. It is based on Martin-Lof Type Theory (MLTT), a family of dependent type theories which have two fundamental features that render them “non-analytic”.
Firstly, there is no notion of an unrestricted variable. Variables are always *typed* which is to say that in order to use a variable, one must always first declare its “range.” This automatically means that any quantification is *restricted*. In first-order logic one starts with unrestricted quantifiers $\forall x, \exists x$ etc. Restricted quantification is then developed in the language of set theory as a derived notion. For $S$ a set, the symbol

$$\forall x \in S(...)$$

is an abbreviation for $\forall x (x \in S \to ...)$. If you think about it for a second, this is grotesque: suppose we want to state a property $P(x)$ about all the natural numbers $x \in \mathbb{N}$. In set theory we do so by saying: “For all things, if that thing is a natural number, then $P$ is true of it.” In symbols: $\forall x (x \in \mathbb{N} \to P(x))$. But unrestricted quantification is not even possible in MLTT. The very syntax of the system forbids it.

The exact details of the system are not relevant here. What is relevant here is that a type theory in the style of MLTT (a dependent type theory) lacks the two disease-bearing syntactic features pointed out above: there is no global equality and there are no unrestricted variables.

13. Now, traditional (extensional) MLTT is still compatible with first-order logic and set theory and indeed can have set-theoretic semantics (i.e. where types are interpreted as sets and terms are interpreted as elements of those sets). So MLTT by itself does not wholly escape the set-theoretic jungle.

But the Univalent Foundations (UF) well and truly do. Their formalizations are based on extensions of MLTT called Homotopy Type Theories (HoTT(s)). There is still a set theory in UF, but UF can no longer take its semantics in set theory: types can no longer be faithfully interpreted as sets. Rather, they must be interpreted as $\infty$-groupoids or,
almost equivalently, *homotopy types*. If we think of these $\infty$-groupoids as “shapes” of a very general kind then we may say that

Univalent Foundations **aim to formalize** *Shape Theory* (or *Homotopy Type Theory* or *$\infty$-groupoid Theory*)

in exactly the same way that

Cantorian Foundations **aim to formalize** *Set Theory*

That is to say, HoTTs are syntactic descriptions of the universe of homotopy types in the same way that ZFC is a syntactic description of the universe of sets. It is important to note, however, that this universe of homotopy types is not to be thought of as “built out” of sets, i.e. they are not to be thought of as “sets with extra structure”. Sets, rather, are to be thought of as built out of them: sets in UF are a special kind of type. Sets in UF are thus a *derived* notion. And so too is first-order logic.

What we have here is a whole new approach to the foundations of mathematics. One in which neither set theory nor first-order logic are primitive notions and which lies entirely outside the spectrum of Cantorian foundations. Thus, one which is *prima facie* incompatible with analytic philosophy.

199. To see this, we must now take a careful look into the *form* of analytic philosophy: problem-solving. To this form we must apply the univalent attitude and see where it leads us. But we must not apply it *in order to solve problems*, but in order to illuminate a new attitude.

What this means: Imagine the current foundations of mathematics were replaced by structuralist foundations based on a dependent type theory. And suppose then that philosophy were to be modelled on those new foundations in the same way that analytic philosophy is now modelled on set-theoretic foundations as expressed in first-order logic.
Then what problems in analytic philosophy would fail to arise in this new univalent philosophy?

When we extract from the Univalent Foundations a philosophical logic to replace first-order logic and set theory, there will be questions that will no longer be well posed. To this I now turn my attention: in particular to Kripke’s skeptical paradox. And the main idea here is that if we can show that the Kripke puzzle can only be expressed because of specific features of the set-theoretic formalism, then we should take that as good evidence that our original question is ill-posed.

67. A dialectical panorama: we have read the history of mathematics into the history of philosophy, and in particular we have identified analytic philosophy as the outgrowth of a particular approach to the foundations of mathematics, best encapsulated by first-order logic and set theory. We have identified a new approach to the foundations of mathematics that gets rid of some of the diseased aspects of set theory. We have identified a formal system that belongs to this new approach. And now we investigate the Kripke puzzle from this new perspective, where we assume that our philosophizing is informed by this formal system (together with its univalent extension) in the same way that canonical analytic philosophy is informed by set theory and first-order logic.

103. Kripke’s skeptical paradox is a kind of limiting puzzle, a problem a discipline faces that is irresolvable without reconsidering its fundamental assumptions or its very form. (This is similar to what Kuhn called a “crisis”, although in philosophy the term must be understood differently, if it makes sense at all.) It is a puzzle that can admit of no solution within analytic philosophy in its current state. It can as much admit a solution as a disease can be a cure for one of its symptoms. But the Kripke puzzle admits of a
resolution, one we may call transdisciplinary: the puzzle fails to carry over to univalent
philosophy, and for purely syntactic reasons.

This “resolution” consists of a seemingly trivial observation: in type theory, there is
no way to even form the question

“Is plus the same as quus?”

when suitably formalized. For what is it that even allows us to pose this question? It is
our conviction that the proposition

\[ + = \text{quus} \]

is well-formed. And that therefore it makes sense to ask whether it or its negation is
ture. This is why Kripke, early on, writes:

[If the Skeptic’s proposal] is false, there must be some fact about my past
usage that can be cited to refute it. **For although the hypothesis**
is wild, it does not seem to be a priori impossible. [...] **[The**
skeptic] questions whether there is any fact that I meant plus, not quus,
that will answer his skeptical challenge.

The Kripke puzzle disappears if we reject this and come to terms with the fact that
the proposition “\(+ = \text{quus}\)” is not well-formed, and that therefore the question is not
well-posed and that therefore the “Skeptic’s” proposal is indeed a priori impossible. To
ask for a fact only makes sense if there is a proposition that that fact could be used to
establish. If there is no such proposition, it is a priori impossible to demand a fact that
establishes it.

The point here is that the Skeptic’s challenge crucially depends on the supposition
that it makes sense to ask of something I was doing in the past, whether **that** thing
was adding or quadding. Namely, that it makes sense to ask for a particular $x$ whether $x = +$ or $x = \text{quus}$. And therefore, that it makes sense to ask whether $+ = \text{quus}$. From a set-theoretic point of view, where everything can be compared to everything else, this seems like an entirely harmless supposition. But it is this very supposition that we must come to reject. To do so, we must abandon first-order logic and set theory as the syntactic underpinning of philosophy. To do so, in short, we must altogether jump ship and abandon analytic philosophy.

The Kripke puzzle seems so difficult from within analytic philosophy because it relies crucially on the principle that any identity statement is well-formed. I should also add: the reason why I chose the Kripke puzzle is because it is a puzzle that arose within and seems specific to analytic philosophy. Of course there are philosophical “puzzles” expressed in terms of identity statements that are far older than the Kripke puzzle (Are mental states the same as physical states? Is moral good the same thing as practical good? etc.) But here I am restricting my attention to a puzzle that arose specifically inside analytic philosophy and for which there exists no historical precedent.

23. Formally speaking, the statement of the Kripke puzzle involves two functions, $+$ and “quus”. We will call the latter $q$. They are functions that seem to take as input two natural numbers $n, m \in \mathbb{N}$ and spit out a natural number as output. These two functions
can be described, informally, as follows:

\[ + : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \]

\[(n, m) \mapsto n + m \]

\[ q : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \]

\[(n, m) \mapsto \begin{cases} 
5 & \text{if } n > 68 \text{ or } m > 57 \\
 n + m & \text{otherwise}
\end{cases} \]

There is something, clearly, that is different about \( q \) and \(+\) simply by virtue of their form. The fact that I need a split brace to define \( q \) and I do not require one to define \(+\) should already make that clear. But the set-theoretic paradigm blurs this formal distinction - in fact, it makes it disappear altogether. They are both functions, therefore sets. We can therefore ask whether they are equal. It is perfectly meaningful to do so.

In type theory, even before we can define \( q \) and \(+\) it becomes clear that they are not terms of the same type. This makes them not formally unequal, but formally incomparable. Indeed we have

\[ + : \mathbb{N} \rightarrow \mathbb{N} \rightarrow \mathbb{N} \]

whereas

\[ q : (\mathbb{N} \rightarrow \mathbb{N} \rightarrow \text{BOOL}) \rightarrow (\mathbb{N} \rightarrow \mathbb{N}) \rightarrow (\mathbb{N} \rightarrow \mathbb{N} \rightarrow \mathbb{N}) \]

These are, to the eye, clearly different types. They are also formally different, which is to say inequivalent types. As such, in type theory it makes no sense to ask whether \( q = + \). The proposition \( q = + \) is not well-formed since \( q \) and \(+\) are terms of different types and the formation rule for the identity type requires as input two terms of the same type.
Therefore the Kripke puzzle is ill-posed: the Skeptic has no right to ask us to compare two incomparable things.

What type theory reveals that set theory obscures is that in $q$ the inputs do not determine the outputs, but rather the function to which the inputs are to be applied. If the inputs are $n = 3$ and $m = 2$ then $q$ instructs us to apply them to the function $+$. If the inputs are $n = 69$ and $m = 52$ then $q$ instructs us to apply them to the constant function $(n, m) \mapsto 5$. By contrast, in $+$ the inputs automatically determine the outputs. This formal difference is laid bare by type theory.

47. “You are comparing a “type” of function (namely a split bracket kind of function) with a “token” function (namely integer addition). Of course their “types” cannot be the same. But what we care about is the token $q$, for which the choice of property (i.e. term of type $\mathbb{N} \to \mathbb{N} \to \text{BOOL}$) has already been made. In particular, by “quus” we mean

$$q_\phi : (\mathbb{N} \to \mathbb{N}) \to (\mathbb{N} \to \mathbb{N} \to \mathbb{N})$$

where $\phi$ is the proposition

$$\phi(n, m) = (n > 68) \lor (m > 57)$$

i.e. $\phi$ is a term of type $\mathbb{N} \to \mathbb{N} \to \text{BOOL}$ given in $\lambda$-calculus notation by $\lambda n.\lambda m.(n > 68) \lor (m > 57)$.”

- $q_\phi$ is still not the same type as $+$ since $q_\phi$ is of type

$$(\mathbb{N} \to \mathbb{N}) \to (\mathbb{N} \to \mathbb{N} \to \mathbb{N})$$
So the proposition-type

\[ q_\phi = + \]

still cannot be formed. So no progress has been made here. Try again.

439. “Fair enough. But then for any given natural numbers \( n, m : \mathbb{N} \) we get that

\[ q_\phi(n, m) : \mathbb{N} \to \mathbb{N} \to \mathbb{N} \]

which has exactly the same type as \( + \). Thus, according to type theory, it is now meaningful to ask whether

\[ q_\phi(n, m) = + \]

And this is exactly the Kripke puzzle. In particular, we can restate the Kripke puzzle as follows: the Skeptic demands of us to prove to him whether we were \( q_\phi(n, m) \)-adding or just plain adding in the past. How do we respond to him?”

- You are right that the types of these two terms are now the same. They are both of type \( \mathbb{N} \to \mathbb{N} \to \mathbb{N} \). But you are not right that the proposition-type \( q_\phi(n, m) = + \) is a statement of the Kripke puzzle. For it is now decidable whether \( + \) is equal to \( q_\phi(n, m) \). It is answerable simply by virtue of the form in which these two terms are expressed. For \( n, m \) here are not arbitrary numbers but specific numbers that have already been chosen. In particular we have that \( + \neq q_\phi(69, 52) \) and \( + = q_\phi(n, m) \) otherwise.

In short, to the question “Are you (or were you) adding or \( q_\phi(n, m) \)-adding?” there is an immediate answer, and it depends on what \( n \) and \( m \) are in the statement of the question. It is impossible to state the question in this form without giving specific values of \( n \) and \( m \). But by so doing, the question automatically
becomes decidable: by merely stating the question, we have given away its answer.

79. “OK, but consider this. Addition itself can be expressed as a term of type

\((\mathbb{N} \to \mathbb{N}) \to (\mathbb{N} \to \mathbb{N} \to \mathbb{N})\)

as follows: we let \(+\) be the function that takes in any two natural numbers \(n, m\) and returns \(+\). \(+\) can thus be thought of as the function constant on \(+\). In \(\lambda\)-calculus notation we have

\[ + = \text{def} \lambda n. \lambda m. + : (\mathbb{N} \to \mathbb{N}) \to (\mathbb{N} \to \mathbb{N}) \]

The proposition-type \(+ = q\) is now well-formed. The Kripke puzzle can now be restated as follows: “Is \(+ = q\)?” And exactly the same issue that Kripke pointed out re-appears.

- But then why stick with \(q\)? Why not just go to \(q\) by adding yet another empty declaration of variables? If I define

\[ +'' = \text{def} \lambda r. \lambda s. \lambda b. \lambda n. \lambda m. + : (\mathbb{N} \to \mathbb{N} \rightarrow \text{BOOL}) \to (\mathbb{N} \to \mathbb{N}) \to (\mathbb{N} \to \mathbb{N} \to \mathbb{N}) \]

then \(+''\) has the same type as \(q\). So there was no need to move to \(q\) to begin with. The same point could have been made with \(q\) since the proposition-type \(+'' = q\) is now well-formed.

This should make it clear what has gone wrong. The problem is that adding a so-called “empty declaration of variables” does not give me back the same thing that I started with! Indeed, the situation is even worse: I cannot even ask whether they are the same: the question itself is not well-formed. For recall what happened: we started with \(+\), and then defined \(+\)', intending for \(+\)' to take
the role of addition. But implicit in this is the assumption that \( + \) is the same thing as \( +' \). But \( + \) is of type \( \mathbb{N} \to \mathbb{N} \to \mathbb{N} \) whereas \( +' \) is of type \( (\mathbb{N} \to \mathbb{N}) \to (\mathbb{N} \to \mathbb{N} \to \mathbb{N}) \). As such, the proposition-type

\[
+ = +'
\]

cannot be formed. We cannot ask the question. You have therefore given me no reason (no expressible reason) to believe that your restated question is the same question as your original question.

167. "OK, but you are merely delaying the inevitable here. What if I define addition as \( +' \)? Then I state the Kripke puzzle as the question: "How do you know that you are \( +'-ing \) or \( q_\phi-ing \)?" Forget about whether this question is formally the same as the original question. Just answer this new question that I've posed for you. There should be a way of telling - when challenged by the Skeptic - whether we have been using \( +' \) or \( q_\phi \). But it seems that we cannot."

- And so we come, inevitably, to the question of naturality (or naturalness.) So I may say here that if it is not impossible to state the Kripke puzzle in a type-theoretic formalism then we may at least confidently say that it is unnatural to do so. For neither \( +' \) nor \( q_\phi \) look anything like addition in the usual sense. But the initial puzzle was supposed to be about following the rules of addition, as those may be taught to a child in elementary school.

The retort here might be that this is exactly what the puzzle is about, namely about knowing what exactly we are teaching a child when we present to it the basic rules of addition. To this, then, I reply that teaching someone (e.g. a child) to \( +'-add \) nor \( q_\phi\)-add involves none of the usual steps that we would use when
we would teach a child. In particular, even in the most trivial case of adding 0 and 1 our explanation would have to involve something radically different. For consider what we would say: “Firstly, pick any two natural numbers $n$ and $m$ - it doesn’t matter which. Then plug them into $+$ to get $+(n, m)$. Now given the type of $+$ we have that $+(n, m)$ is of type $\mathbb{N} \to \mathbb{N} \to \mathbb{N}$. To this you must then plug in 0 and 1, and voila, this is how you $+$-add 0 and 1.” Now one might imagine a child being lost even at the first step (What does “arbitrary” mean - What are “natural numbers”?) but even someone familiar with these concepts will surely immediately know that something has gone wrong. If he is meant to be learning about an operation on the natural numbers, then why is he being taught to make some arbitrary choice in the beginning? It is unclear to him what we are trying to teach him, but it is perfectly clear to him that what we are trying to teach him to do is not to apply a binary operation on the natural numbers.

223. “So what? The point is not to stay close to the intuitive notion of addition. The point is to be able to state something analogous to the Kripke puzzle in this fancy new type-theoretic system of yours. And we’ve managed to do just that, since $+$ = $q_\phi$ is a well-formed proposition-type. Even if the process of teaching someone how to $+$-add and how to $q_\phi$-add are very different from the process of teaching someone how to add (in particular the latter won’t involve teaching someone to make some arbitrary choice as a first step, for example), what matters is that the process of teaching one is very similar to the process of teaching the other. In fact they are seemingly identical and that is why the Kripke puzzle re-arises: for someone who has been taught how to $+$-add for the first few examples (e.g. for small numbers) will not be able to tell if he had been $+$-adding or
$q_\phi$-adding if so pressed by the Skeptic at some later time. And type theory says nothing about that difference.”

- I disagree. The process of teaching someone how to $+'$-add will differ from the process of teaching someone how to $q_\phi$-add, not at the first step but at the second step. The first step of $+'$-adding involves making an arbitrary choice, i.e. picking any two natural numbers. But the first step of $q_\phi$-adding the choice is not arbitrary: one chooses the numbers that one eventually wants to apply $q_\phi$ to. So if we want to $q_\phi$-add $0$ and $1$ what we do is we first test $\phi(0, 1)$ to see if it is true or false and then return $5$ or $0 + 1$ accordingly (in this case since $0 < 68$ and $1 < 57$ we will return $0 + 1$. But in the case of $+'$ we are not required to pick the same numbers that we want to $+'$-add in that first step. Indeed we have

$$+'(3, 300)(0, 1) = +'(0, 1)(0, 1)$$

It doesn’t matter at all what choice we make at the first step.

359. “You are nit-picking. Since it makes no difference, then let’s just assume that the way you teach someone to $+'$-add is that the first step is to plug in the numbers that they are given (rather than, say, arbitrary ones like 3 and 300.) Then that first step in learning how to $+'$-add becomes the same as the first step in learning how to $q_\phi$-add. And then the argument goes through.”

- Yes, indeed. But the second step is not the same - and it can never be. For the second step of $q_\phi$-adding, say, $0$ and $1$ involves testing the truth or falsity of $\phi(0, 1)$. But the second step of $+'$-adding is simply to evaluate $0 + 1$. Even in the case of $0$ and $1$ the process of teaching someone how to $+'$-add or $q_\phi$-add diverge. And dealing with this last objection reveals to some extent the principle of this
entire tortuous dialogue: in type theory one cannot define a function unless one gives a process with which one evaluates it (this has to do with the fact that MLTT was originally meant as a foundation for constructive mathematics.)

439. “But now you are outright cheating! If such a move were allowed then the original Kripke puzzle becomes trivial, since the Skeptic can be immediately rebuffed: ‘If I had been taught how to quadd instead of how to add, then I would remember my teacher explicitly teaching me how to do something that is not taught when I learn how to add.’ Namely, I’ll be able to point to the following difference between quus and plus: In teaching someone how to quadd, you must also teach them to test their inputs on what you have been calling φ. But such a move cannot be allowed, in fact altogether defeats the whole point of the puzzle, which assumes that the first steps of teaching someone how to quadd are the same as those carried out when teaching someone how to add. For adding clearly involves no such test and so one would immediately know if they had been taught to add or quadd (simply by checking if their instructor had mentioned φ or not.)”

• Yes, cheating. You are right: to you it must appear like cheating. But so too is (as far as you are concerned) the restriction type theory imposes on the well-formedness of identity statements. So too is forbidding the notion of a “wholly undetermined” variable. These are “cheats”, in the sense that we are no longer trying to outmanoeuvre the puzzle, but rather to redesign our racing engine from scratch. But they are “cheats” implemented with a specific reason in mind - and that is what makes them permissible.

Once again the whole principle of the “argument” is revealed here: in type theory it is less important to prove that you have “grasped a particular rule” than it is to be clear on what kind of rule you are trying to grasp. And usually
when you are fully explicit about the kind of rule you are trying to grasp (or learn, or teach) no ambiguity remains as to whether you’ve grasped this or that rule. I should also remind you that the problem becoming trivial is exactly my desired conclusion: so I completely agree with your assessment. It seems to me that we have reached a kind of agreement: in type theory Kripke-style problems do not arise except as trivialities.

7. “We have not reached any kind of agreement! There are still so many ways to restate Kripke’s puzzle in your formal set-up! Here’s the most obvious one: It is meaningful in the kind of type theories you want to consider (type theories with “type universes” $\mathcal{U}$) to ask whether a given type $A$ is equal to another type $B$. You do so by forming the identity type $A =_{\mathcal{U}} B$ over $\mathcal{U}$. Therefore it is meaningful - by your lights - given terms $a: A$ and $b: B$, to ask whether $a$ is of the same type as $b$. Replacing $A$ with

$$\left(\mathbb{N} \to \mathbb{N} \to \texttt{BOOL}\right) \to \left(\mathbb{N} \to \mathbb{N}\right) \to \left(\mathbb{N} \to \mathbb{N} \to \mathbb{N}\right)$$

and $B$ with

$$\mathbb{N} \to \mathbb{N} \to \mathbb{N}$$

and $a$ with $q$ and $b$ with $+$, we get that it is meaningful to ask whether $q$ is of the same type as $+$. And then there seems to be no fact of the matter to determine whether you were applying an operation of type $A$ or an operation of type $B$, etc. etc.”

- No. Here you are simply misunderstanding the bigger picture. The point is not to restate a form of the Kripke puzzle based on this new formalism. Of course this is formally possible - indeed there is even a “set theory” inside that new formalism and you can just take it from there without any need to talk about “type universes” and what have you. The point isn’t even to show that it is
unnatural to state it in such a formalism - that would push us into an irrelevant debate.

The point, rather, is this: Kripkenstein-style considerations on rule-following altogether ignore the idea that before one can even start learning how to follow a rule, one must be fully explicit about the kind of rule that one is about to try to learn how to follow: “I am going to teach you how to follow a rule, and this rule is of a certain kind $X$” - the latter half of the sentence seems completely irrelevant. Our intuition here is probably that in learning how to apply a rule, information about what kind of rule it is that we are to be taught to apply is redundant. Such information might come in handy later - e.g. for taxonomic reasons - but it is not of the essence when we are still learning.

Such an attitude appears perfectly reasonable in a set-theoretic mindset. For example, in set theory there is no apparent difference between a function on the natural numbers defined recursively or defined using a split brace, or defined using this or that method: they are all functions and they can all be compared to each other. So it would appear completely irrelevant to say: “Now I am going to teach you how to add, where addition is a recursive function on pairs of natural numbers etc. etc.” (But notice how the Kripke puzzle could not even be stated if this is how people were taught how to add or quadd.)

Such an attitude becomes unreasonable – even impossible and absurd – in a type-theoretic mindset. Even more is true: in a type-theoretic mindset any ambiguity as to which rule I seem to have grasped altogether disappears, as long as the type of the rule is correctly and explicitly specified. This is what my “argument” is about: not about being unable to use this new formalism to state some or other variant of the original Kripke puzzle.
191. We have reached our desired conclusion: distinguishing between plus and quus is just not a meaningful distinction to make in univalent philosophy. As such, a puzzle that involves asking whether plus and quus could be the same - or, equivalently, how we can know that they are different - cannot be stated in a univalent philosophy.

But the moral I urge we draw from this “solution” – as indicated by the last answer – is that we should entirely move away from the idea of “solving problems” in philosophy. To view it as a solution in the usual sense altogether defeats the point of what I am trying to do here.

What I am trying to say boils down to this: the style of argument that can be applied to resolve the Kripke puzzle is unrecognizable as analytic philosophy – indeed utter garbage from the point of view of those who consider it a real problem. This is exactly the sense in which it is a limiting puzzle. So even though my “argument” here can be read as a traditional argument within analytic philosophy (there is a thesis, there are premises, there are inferences and conclusions, characters and dialogue) that is not my intention here at all. In particular, the result of the “argument” is not to be understood as the assertion or negation of a particular thesis – it is rather to be thought of as a pastiche, a feigned attempt at logical argumentation, indeed a kind of joke, a gaiety. The real underlying argument consists in the very form of this Appendix (in which the Kripke “argument” is embedded as one of its movements.)

139. What I am trying to do here is not to advocate a new formal system, to replace the old. I am not saying we should adopt a different logic, for instance. That too is unacceptable to me. Nor am I saying that we should look at mathematical proofs or scientific experiments and from them determine that a particular philosophical question
has this or that answer (or no answer). There is an element of false prophecy in doing this, and there are already too many false prophets in analytic philosophy.

To say that the Kripke puzzle cannot be stated in a univalent philosophy is neither to solve it nor to dissolve it. It is merely to point out that puzzles of that form are not stateable when one enters a type-theoretic mindset. But even more is true: the dagger must be pushed even deeper. In univalent philosophy one should no longer think of the form of the discourse as consisting of statements of “puzzles” made precise by “definitions” and then resolved by “arguments”. The very idea that one is doing philosophy when one is stating and tackling the Kripke puzzle must be called into question.

But crucially: not in order to claim that we have now discovered the true form of philosophy. Such an attitude itself is of the “definitions-puzzle-argument” variety, i.e. based on the idea that there is progress in philosophy.

59. Wittgenstein on civilization:

Our civilization is characterized by the word ‘progress’. Progress is its form rather than making progress being one of its features. (Culture and Value)

Exactly analogously, problem-solving is the form of analytic philosophy rather than one of its pleasant features. Anyone who has been fascinated by one of its problems must have felt this at one point or another. And this form is again probably due to Russell. In “On Denoting” he writes:

A logical theory may be tested by its capacity for dealing with puzzles, and it is a wholesome plan, in thinking about logic, to stock the mind with as many puzzles as possible, since these serve much the same purpose as is served by experiments in physical science.
The conception of philosophy as consisting of the assertion or negation of “logical theories” whose adequacy boils down to its capacity to “deal with puzzles” is something entirely new – and entirely artificial.

Problem-solving will not be the form of univalent philosophy. The form of univalent philosophy - consistent with the secret history of philosophy - will resemble the new style of doing mathematics.

31. The old style of mathematics – Definition, Theorem, Proof – if it ever existed (because logic has certainly twisted our views on this) is now obsolete. The new style of mathematics is this: Perspective, Structure, Construction. You begin by re-arranging existing data into a new perspective, which may initially have nothing to do with your existing data and which need not be justified by it at all. You then put structure on that new perspective - structure that may initially have absolutely nothing to do with your original data, or its original purpose. And finally, using that structure, you construct what you need. This is exactly the kind of attitude, for instance, that allows contemporary mathematics to study the collection of all geometric objects of a certain kind as a geometric object in its own right. (The collection of all topological spaces as a “generalized space”.) And there are already examples of actual mathematics being produced in something approaching this style. (The old style of doing mathematics can be recovered, of course, as a special case: Definitions are the least ambiguous form of Perspective; Theorems the most concrete case of Structures; Proofs the most rigid forms of Construction.)

Ulam’s spiral serves us here as a methodological prototype, a blueprint for the transition from the analytic style to the algebraic or synthetic or type-theoretic or univalent style. Philosophy must now forget the old style of mathematics that it has copied and
move to the new one. There is nothing universal or sacred about this new style, of course. It will not solve problems - problem-solving is the old way of understanding what constitutes progress: we cannot try to understand the new style through the old one: it would be like trying to see through new glasses while keeping our old contact lenses on.

This new style will not lead us anywhere. We cannot be led anywhere by anything. But it will be just one more step, just another vocabulary that we must use, abuse and discard before...Before what? “Before” implies an after, and that is not the point at all, because we are once again thinking in terms of progress.

The point is this: I am not presenting this as a first step of some “research program” that would, say, (re-)examine arguments in analytic philosophy from a univalent point of view in order to, say, solve, resolve, absolve or dissolve them. What I am saying, instead, is that univalent philosophy would be a philosophy that entirely moves away from the idea of philosophy as a discipline dedicated to the “examination of logical problems and puzzles.” It is also a philosophy that would most emphatically reject the absurd notion that there are “ahistorical problems” that philosophers have tackled through the ages - a conviction that we must surely attribute at least partly to the diseased idea that since all of mathematics can be formalized in one unique language, all mathematicians are thinking of their problems in one unique shared eternal and ahistorical language. What perhaps is not clear to philosophers now (and which this work aims at revealing) is that such a philosophy can also be based on a formal system just as rigorous as first-order logic and set theory. This, I think, is probably still hard to imagine.

127. This “Appendix” does not merely contain philosophical or mathematical content. It also mimics their form. The content (if I can even call it such) reveals itself primarily in
the form in which it is expressed: a spiral punctuated with all the transversal movements cutting across and through it.

So I urge you to think of the content of this “Appendix” as follows: Several threads of observation, aphorism, exposition and argument have been arranged in the form of a spiral: Perspective. The numbering reveals the transversal movements on that spiral: Structure. Its use is in illuminating a manner in which an argument can be constructed that altogether avoids the “Definition-Puzzles-Argument” co-ordinate system: Construction.

And exactly the same kind of thinking must be applied to the history of philosophy, as it relates to what is expressed here. View the history of philosophy as mirroring the history of mathematics, and especially the history of mathematics as it moves through cycles of rigour and unrigour: Perspective. The idea of normal philosophy and transcendental philosophy can then be represented by tangent or intersecting lines through this spiral: Structure.

As for Construction: See me as advocating a new form of argument not by explaining it, but by mimicking it in order to demonstrate, using a synthesis of old-style and new-style philosophical argument that it is possible to model philosophy after a formal system without making philosophy a game of formal deductions. How exactly all these spirals – secret, fictional, or otherwise – relate to each other and to the points being made here, this I leave to you to figure out, starting from the observation that all paragraphs in here have been numbered by prime numbers.

7. Here is an example of an irrelevant distinction that set theory allows us to make:

The proposition

\[ 1 \in \mathbb{Z} \]
is a well-formed sentence in the language of set theory. It is true of \( \mathbb{Z} \) but it is not true of \( 2\mathbb{Z} \), even though \( \mathbb{Z} \) is isomorphic to \( 2\mathbb{Z} \) as a group.

More precisely, the set of groups is definable in set theory (e.g. ZFC) by a one place open formula \( \gamma \) in the language of set theory. This means that there is a set \( G = \{ x | \gamma(x) \} \) such that

\[
x \in G \iff x \text{ has the structure of a group}
\]

In particular, \( \text{ZFC} \models \gamma(\mathbb{Z}) \) and \( \text{ZFC} \models \gamma(2\mathbb{Z}) \). But we also have

\[
\text{ZFC} \models \phi(\mathbb{Z})
\]

whereas

\[
\text{ZFC} \not\models \phi(2\mathbb{Z})
\]

Even more odiously ZFC also defines when two groups are isomorphic, i.e. there is a two-place open formula \( \iota \) such that

\[
\text{ZFC} \models \iota(G,H) \iff G \text{ and } H \text{ are groups and are isomorphic}
\]

So in particular we get

\[
\text{ZFC} \models \iota(\mathbb{Z}, 2\mathbb{Z}) \land \phi(\mathbb{Z}) \land \lnot \phi(2\mathbb{Z})
\]

This statement above, if you think about it, is an abomination and even I would say conceptually a contradiction: it expresses that two things are indistinguishable and there is something that is true of one of them that is not true of the other. It is a blatant violation of Leibniz’s law. The “solution” to the contradiction of course is that the notion of “indistinguishability” being expressed by \( \iota \) is coarser than extensionality of sets and so set theory is perfectly within its rights to assert that two things are structurally
indistinguishable yet extensionally unequal. But to be able to do that is not a feature, it’s a bug (to reverse an oft-quoted observation of computer scientists.) In particular it allows us to ask such questions as: “Is π equal to the set of prime numbers?” That is to say, it allows that such questions are grammatical.

13. An ∞-groupoid can be thought of as a general shape. Or rather, as encoding the general “type” of an abstract shape. More precisely, an ∞-groupoid corresponds to a topological space, up to homotopy.

What does this mean, intuitively? What is the “homotopy type” of a particular topological space? Roughly, it is the shape of that space insofar as it can be measured by tying pieces of string around it. In particular, we are interested in ways of wrapping the string around a shape such that there is no way to remove the string without cutting it.

Consider a sphere and a donut:

Any string we tie around a sphere can be removed from the sphere without cutting it, simply by “sliding” it out. But tie a string around the “tube” of a donut and there is no way to remove it without cutting it. Any way we “slide” it, it will remain wrapped around the tube. This indicates that the sphere and the donut are not the same “up to homotopy”. They have different homotopy types: they have different shapes.

A homotopy type is a combinatorial representation of these shapes in terms or nodes, arrows between nodes, arrow between arrows between nodes, etc. etc. all the way up to infinity.
31. We have, for example, the following case of Perspective-Structure-Construction:

\[
\begin{array}{cccccccc}
37 & 36 & 35 & 34 & 33 & 32 & 31 \\
38 & 17 & 16 & 15 & 14 & 13 & 30 \\
39 & 18 & 5 & 4 & 3 & 12 & 29 \\
40 & 19 & 6 & 1 & 2 & 11 & 28 \\
41 & 20 & 7 & 8 & 9 & 10 & 27 \\
42 & 21 & 22 & 23 & 24 & 25 & 26 \\
43 & 44 & 45 & 46 & 47 & 48 & 49 & \ldots
\end{array}
\]

**Figure 1.** Perspective: Positive natural numbers arranged in a spiral
Figure 2. **Structure**: Prime numbers form diagonals cutting across the spiral

**Construction**: Looking at the kind of structure that emerges from this new perspective, i.e. the fact that there are diagonals cutting across the spiral that seem to be dense in primes, we realize that it makes sense to try and construct polynomials that tend to produce an unexpectedly large number of prime numbers, e.g. “Euler’s prime generating polynomial” \( x^2 - x + 41 \).

73. The set-theoretic style has infected both analytic and continental philosophy. One finds, for instance, both Kit Fine and Alain Badiou arguing about the ontological status of the empty set. And there are, I’m sure, people who read one of them seriously who don’t even know who the other is.

I think both Fine and Badiou understand set theory. I’m not criticizing them for misunderstanding set theory; indeed I am not criticizing them at all. I am merely pointing out that they have both adopted set-theoretic language as an adequate language in which to state, argue and develop problematics in metaphysics (and specifically ontology).

There is more to admire than to deride in such projects. (Enough of the false prophets of the sciences or of mathematics (e.g. Sokal) thinking they can prove that such-and-such a formal concept has been misunderstood by a philosopher. Off you go, back to
Philosophers are certainly often unclear about how empirical science relates to and supports their work, and do occasionally become seduced by the idea of a “decisive refutation” of a position through some formal result (Russell’s paradox is here the prototype.) But that kind of thing happens all the time even in science.

What I am trying to clarify here is this: I am not arguing that set theory and first-order logic are being misused and that perhaps they could be used to do philosophy, if only people would do it the right way. Quite the opposite. The point is to change the formalism altogether, and then imitate this new formalism in the same way that analytic philosophy (and occasionally even continental philosophy) has been imitating first-order logic and set theory. What exactly this means depends on form as well as content.
CHAPTER 2

Univalent Foundations as Structuralist Foundations

In one way or another, structuralism in mathematics is about getting rid of inessential properties of mathematical objects. It doesn’t matter whether the number 3 is written as III, iii or \{\{\emptyset\}\}. What matters is that 3 comes before 4 and after 2. It doesn’t matter whether 3 contains \{\emptyset\} as an element. What matters is that 3 is the immediate successor of the only even prime. Thus, on some level, to be a structuralist about mathematical objects is to believe that the only properties of those objects that matter are those we may call structural properties. Everything else we can say of mathematical objects is nonsense (or, more charitably, non-mathematical).

What could this possibly have to do with the foundations of mathematics? Here is a wild thought: What if the non-structural properties of mathematical objects could be ruled out purely by grammatical considerations? In other words, what if expressing mathematical nonsense like “\{\emptyset\} ∈ 3” became a grammatical impossibility rather than a philosophical curiosity? A structuralist foundation for mathematics – as I will use the term – would be a formal system that achieves this, i.e. one in which to express nonsense becomes a grammatical impossibility. In this Chapter, I will argue that the Univalent Foundations of mathematics provide a foundation that does justice to this idea.

Other than perhaps the early Wittgenstein, the first person to envision such a “nonsense-free” foundation was M. Makkai in his writings on First Order Logic with Dependent Sorts (FOLDS) [110, 111, 113]. Needless to say, an important precursor is the work of Lawvere [95] on the Elementary Theory of the Category of Sets (ETCS) which forms
a prototypical example of what has come to be known as a “categorical” (or “category-theoretic”) foundation of mathematics, as well as observations of Freyd [46] and Blanc [23] on how dependently-typed languages can be used to avoid making non-structural distinctions.\(^1\) Despite claims made implicitly by Lawvere [95] and explicitly by McLarty [131, 132] (among others) to the effect that ETCS-style foundations are compatible with a broadly structuralist view of mathematical objects, there was no attempt to systematically restrict the language of these foundations so that it is able to express only structural properties. In this endeavour, Makkai was undoubtedly the pioneer.\(^2\)

Makkai’s vision for a “nonsense-free” foundation of mathematics inspired V. Voevodsky to develop his own new foundational proposal going under the name of *Univalent Foundations* (UF) [171, 175, 177, 180]:\(^3\)

Univalent Foundations can be seen as a realization of the vision of Michael Makkai [which] was very important for me in my search for a formal language for contemporary mathematics. ([178], p. 2)

In realizing this vision, UF has exploited a fascinating connection between abstract homotopy theory and Martin-Löf Type Theory (MLTT).\(^4\) The technical development underlying UF has come to be known as Homotopy Type Theory (HoTT) [171]. Very roughly, HoTT consists of the idea that the basic objects of MLTT (“types”) can be interpreted as homotopy types (“abstract shapes”). The key feature of HoTT/UF – and what connects

\(^1\)ETCS is also called a “structural” set theory, as opposed to an “extensional” or “material” one like ZFC. For accessible comparisons of the two see [96, 143].

\(^2\)Although I will only be dealing directly with Makkai’s system in Chapter 3, let me mention that in its most recent incarnation in [113] it goes under the name of Type-Theoretic Categorical Foundations of Mathematics (TTCFM). See Marquis [122] for an exposition as well as philosophical defence of Makkai’s views on the foundations of mathematics.

\(^3\)It is important to note, however, that Voevodsky’s and Makkai’s motivations for their respective projects are very different. Makkai’s motivation was at least partly philosophical, aiming to develop a language that encodes a structuralist view of mathematical objects. On the other hand, Voevodsky’s work on foundations had primarily a non-philosophical motivation behind it: to develop usable “proof assistants” that mathematicians can use to verify their theorems.

\(^4\)This connection was discovered independently by Voevodsky [175] and by Awodey and Warren [11, 183].
it both to structuralism in the philosophy of mathematics and to “nonsense-free” foundations in Makkai’s vein – is the so-called *axiom of univalence* which states, roughly, that isomorphic objects are identical.\(^5\) Awodey [9] has already made a start in highlighting the connections between UF, univalence and structuralism and in this Chapter I will sharpen and further reinforce them. In particular, I will argue that UF is better able to do justice to the ideal of a structuralist foundation than any other foundational proposal currently available, at least for a particular way of making precise what this “structuralist ideal” is supposed to be.

In Section 2.1 I will outline a minimalist conception of what constitutes a foundation of mathematics and explain how the Univalent Foundations fit into this picture. In Section 2.2 I will offer a pragmatist reading of the structuralist thesis in the philosophy of mathematics and then explain how it ought to apply to the foundations of mathematics. I will then articulate a precise criterion – (SFOM) – that a foundational system ought to satisfy if it is to be called structuralist in my sense. I will then examine the extent to which ZFC and ETCS fail (SFOM) (Section 2.2.1) and explain how (SFOM) shifts the focus from ontology to language, thereby transcending the debate about the foundational fitness of set theory vs. category theory (Section 2.2.2). In Section 2.3, I will argue that the axiom of univalence ensures that UF satisfies the “principle of isomorphism” which entails, roughly, that isomorphic objects are identical. I will spell out two further criteria that, if true, would ensure that UF comes closest to (SFOM) than other proposals. In Section 2.4 I argue that the second criterion is satisfied by describing a general method for formalizing criteria of identity as homotopy equivalences. Finally, in Section 2.5, I will argue in favour of the first criterion by responding to several general challenges that

\(^5\)A more correct way of paraphrasing the axiom of univalence is this: *identity is isomorphic to isomorphism*. A fully correct way of paraphrasing the axiom of univalence is this: *the (canonical) map that sends identities to isomorphisms is an isomorphism.*
have been put forward against the foundational aspirations of UF and by examining the scope of the general method of Section 2.4.6

2.1. Univalent Foundations, HoTT and intensional MLTT

I take a foundation of mathematics (or foundational system) to consist of the following elements:

\[
\text{Basic Language} \xrightarrow{\text{used to express a...}} \text{Theory} \xrightarrow{\text{that describes a...}} \text{Universe of Objects}
\]

where the “Basic Language” is a formal language of some sort, the “Theory” is a formal theory establishing a standard of rigour (usually through a notion of formal deduction) and the “Universe of Objects” is some intuitively comprehensible collection of objects such that (there is good evidence that) all of currently practiced mathematics can be encoded in terms of them. Clearly, this is intended to be a very minimalist conception of a foundation of mathematics. One will usually want to impose extra conditions, e.g. that the “Basic Language” is finite or that the “Theory” is consistent or that the “Universe of Objects” is philosophically coherent in some appropriate sense. I will here refrain from doing so explicitly – this adds generality to the theses that I will argue for, but also, more importantly, indicates that the demand of (SFOM) that I will state in Section 2.2 below does not depend on any notion of a foundation that goes beyond this minimalist conception. From now on, unless otherwise specified, by a foundation for mathematics I will mean any collection of components that fit into this minimalist picture.

---

6An early version of this chapter was presented in an APA special session on Univalent Foundations in Philadelphia, December 2014. A slightly altered version of this Chapter has been accepted in Synthese as [169].
For **set-theoretic foundations** perhaps the most widely accepted set-up in terms of the above picture is the following

\[
\text{First-Order Logic with equality} \quad \xrightarrow{\text{...used to express...}} \quad \text{ZFC} \quad \xrightarrow{\text{...that describes...}} \quad V
\]

where \( V \) is the “real” cumulative hierarchy of sets. Of course, within set-theoretic foundations both (1) and (2) are subject to variation. For example, (1) could be replaced by a Russellian type-theoretic syntax instead of first-order logic (in which case (2) would, say, consist of asserting the axiom of reducibility among other axioms). Or we could maintain first-order logic as our (1) but use it to axiomatize different set theories in (2), e.g. ZF or intuitionistic ZF.\(^7\) Furthermore, one could argue philosophically about the metaphysics of (3), i.e. the “real nature” of \( V \). What seemed inconceivable until very recently was a foundational system (in the above sense) in which component (3) was filled out by anything other than a universe of sets. To be sure, there was disagreement about what that universe of sets *really* should be. For example, should it understood as an ETCS-style category or a ZFC-style cumulative hierarchy? But hardly anyone imagined the possibility of a foundation of mathematics in which the universe of basic objects was not meant to be understood as a collection/hierarchy/universe of sets of some kind.

With the **Univalent Foundations** of mathematics (UF), this becomes a real possibility. UF is a foundational proposal that differs from the standard set-theoretic set-up in *all* of the above components (1),(2) and (3). The standard set-up for UF – and the

---

\(^7\)I am assuming here that the *deductive system* that we impose on the basic formal language comes in at stage (2) – but this is admittedly a difficult assumption to maintain in the case of MLTT where the rules of well-formedness of expressions coincide with the deductive rules. Still, even in dependent type theories used to formalize UF, there is a useful distinction to be made between “basic” type-formers (e.g. \( \Pi \), \( \Sigma \) or \( W \)-types) and type-formers or axioms added specifically with the homotopy interpretation in mind.
one I will consider for the purposes of this Chapter – is as follows

\[
\text{Intensional MLTT} \quad \vdash \text{HoTT} \quad \vdash \infty\text{-Grpd}
\]

where \(\infty\text{-Grpd}\) is the “real” universe of homotopy types. Thus, by the term \textbf{Univalent Foundations} I will refer to a new foundation of mathematics of which HoTT is just one of many possible formalizations and intensional MLTT one of many possible basic languages.\(^8\) I will now briefly introduce the above-stated components (1),(2) and (3) in the case of UF.

“Intensional MLTT” refers to the cluster of formal systems that usually go under the banner of \textit{intensional Martin-Löf Type Theory} (MLTT) – for classical introductions see the original paper \cite{124} as well as expositions in \cite{67,168}. MLTT was originally an attempt to give a constructivist\(^9\) and computer-friendly foundation of mathematics. It comes in \textit{extensional} and \textit{intensional} variants and we are here concerned exclusively with the latter. As a formal system, the basic objects of MLTT are \textit{types} and \textit{terms}.\(^{10}\)

A \textbf{judgment} is a syntactic expression asserting a fact about terms and types in a certain \textbf{context}. The rules of the system tell us how to construct new types and terms from old ones using various type constructors, including but not limited to the \textbf{empty type} \((0)\), the \textbf{singleton type} \((1)\), \textbf{product types}, \(\Pi\)- and \(\Sigma\)-\textbf{types} etc.\(^{11}\) There are two main innovations in MLTT that set it apart from the “type theories” that preceded it,

\(^8\)That said, I will often use the terms UF and HoTT interchangeably, i.e. I will often refer by the term UF to the \textit{particular} formalization of UF as HoTT. This is in keeping with standard set-theoretic practice of referring to \(\text{ZFC}\) as “set theory”.

\(^9\)Martin-Löf’s original name for his theory in \cite{124} was \textit{Intuitionistic Type Theory} - we shall not be concerned here with terminological issues concerning the difference between “constructivists” and “intuitionists”. Sufficient to say that although MLTT has most often been associated with constructive/intuitionistic approaches to mathematics, UF is in no way limited to constructive/intuitionistic logic. Indeed, HoTT is perfectly consistent with an appropriately stated law of the excluded middle (cf. \cite{171}, Definition 3.4).

\(^{10}\)One can think intuitively of types as sets and terms as elements of sets although in HoTT this way of thinking is misleading since one can prove that there are types that are not \(h\)-sets (cf. \cite{171}, Example 3.1.9).

\(^{11}\)For a complete list of rules for a system of HoTT see the Appendix to Chapter 3.
e.g. Russell’s and Church’s. Firstly, MLTT introduces the notion of a dependent type. The idea is that we may “index” types by terms of another type - in which case we say that the family of types being indexed depends on the indexing type. As a formal judgement this would be written as \( \Gamma, x : A \vdash B(x) \) \textbf{Type} which is to be read as follows: “In context \( \Gamma \), given a term of type \( x : A \) we can produce a type \( B(x) \).” Thus we say that the type \( B \) depends on \( A \). Secondly, MLTT introduces the notion of an identity type. As with other type constructors the identity type is governed by four deductive rules: formation, introduction, elimination and computation. The so-called formation rule, for example, says that given any type \( A \) and any terms \( a,b \) of type \( A \), it is possible to form a type \( a =_A b \), to be thought of as the type of “proofs of identity” between \( a \) and \( b \).

By “Homotopy Type Theory” (HoTT) I will mean the formal system laid out in the Appendix of [171], namely intensional MLTT together with new deductive rules governing univalent universes and higher inductive types. A univalent universe is a universe of types (think: “class of sets”) \( \mathcal{U} \) satisfying the univalence axiom which states that the identity type between types is equivalent to homotopy equivalence. In HoTT the types of MLTT are no longer to be understood as sets, but rather as homotopy types or \( \infty \)-groupoids. This claim has been made precise in [77], where a model of HoTT is constructed in which types are interpreted as Kan complexes, which provide a widely accepted model for \( \infty \)-groupoids constructed set-theoretically.

---

\(^{12}\)It is important to note that this particular extension of MLTT is not the only homotopy type theory that formalizes UF. A recent alternative proposal is Cubical Type Theory [21, 31] designed with the specific intention of giving a constructive model of univalence. Nevertheless, for the purposes of this Chapter, nothing hinges on the choice of formalization and so we will identify the term HoTT with the formal system used in [171].

\(^{13}\)For the purposes of this Chapter these two terms will be used interchangeably. However, the correspondence between homotopy types and \( \infty \)-groupoids is usually regarded as a non-trivial hypothesis (the “Homotopy Hypothesis”) which can be made precise and proven in various ways.
Finally “$\infty\text{Gpd}$” denotes the universe of $\infty$-groupoids. $\infty$-groupoids are algebraic objects meant to capture the homotopy types of topological spaces. This correspondence between $\infty$-groupoids and homotopy types is called the “Homotopy Hypothesis” and there are various methods of making it precise. Perhaps the most widespread such method involves defining $\infty$-groupoids as Kan complexes in the category of simplicial sets (cf. [70, 103] as well as the Appendix to Chapter 4) with its standard model structure (cf. [70, 149]). But surely if UF is to be regarded as a self-standing foundation we cannot rely on the definition of $\infty$-groupoids as Kan complexes inside ZFC. Nor should we rely on the mathematical definition of the homotopy group of a topological space, especially if that topological space is understood as a set of points with a topology. In other words: is there a “naive” way of understanding homotopy types? Broadly, yes: homotopy types are to be thought of as abstract shapes, just as sets are to be thought of as (abstract) collections. Why the qualifier “abstract”? Because homotopy types are – as the term itself suggests – meant to encode types of shapes that share some common structure. The homotopy type itself is the “skeleton” that is common to all particulars belonging to the given type – hence an abstract shape. But this does not mean that these “skeletons” are not amenable to concrete representations. In fact, the theory of $\infty$-groupoids emerged partly as an attempt to represent them as concrete combinatorial objects. What was scarcely imaginable was that these objects could then be employed as the basic objects of a foundations of mathematics, axiomatizable by a formal system in very much the same way that sets are. But it is this very possibility that allows us to consider defining structuralist foundations in the sense that I will now outline.

\[^{14}\text{For a more detailed exposition of a similar kind of view of homotopy types, as well as an argument for their fundamentality, see [121]. A pre-formal account along those lines will be taken up in Chapter 4.}\]
2.2. Structuralist Foundations of Mathematics

Following Benacerraf [19], philosophers advocating a structuralist position in the philosophy of mathematics take themselves to be arguing in favour of some reasonable variation of what we may call the structuralist thesis: Mathematical properties of mathematical objects are concerned with the relations these objects bear to each other, rather than what these objects are. Parsons [145], for example, expresses this as “the view that reference to mathematical objects is always in the context of some background structure, and that the objects involved have no more to them than can be expressed in terms of the basic relations of the structure” and Resnik [151] as the view that “the objects of mathematics [...] are structureless points or positions in structures [that] have no identity or features outside of a structure.”

Rather ironically, several disputed structuralist positions are then carved out through ontological disagreements on what “structure” itself is. There is, for instance, disagreement over whether structures are possibilia (Hellman [61] and, more obliquely, Putnam [148]) or Platonic natureless abstractions (Shapiro’s ante rem structuralism in [162]) or even just sets (following in the Bourbakian tradition). This direction takes the structuralist thesis to be an observation about ontology: mathematical objects are structures. Put differently: what it is for something to be a structural property is to be determined by figuring out what structure means.

On the other hand we may take the structuralist thesis to be an observation about mathematical practice. Namely: the only mathematical properties of a mathematical object that are meaningful in practice are those that are invariant under “isomorphism”, where “isomorphism” is a place-holder for the relevant criterion of identity for the type of object under consideration and where what is “relevant” is to be understood by recourse to mathematical practice. For example, the only meaningful mathematical properties of
The mathematical properties of dense linear orders are those that are invariant under order-preserving bijections. Group isomorphism and order-bijection are relevant criteria of identity because practicing group theorists or number theorists (working, say, over $\mathbb{Q}$) would agree that they preserve the truth of the statements they are interested in. In short, we may define as structural those properties that mathematicians themselves would find meaningful (or “sensible”). Let me call this the practical reading of the structuralist thesis.

There are two ways to make good on the practical reading of the structuralist thesis. Firstly, informally. This is roughly the direction taken by Awodey [6] and McLarty [131, 132] (“schematic” or “categorical” structuralism) and more recently by Burgess [26, 27] (“permanent parameter” structuralism). Secondly, formally – this is where my interests lie. The focus here is more on the foundations of mathematics rather than on the philosophy of mathematics. Specifically, we take the practical reading of the structuralist thesis to indicate a “design constraint” for a foundation of mathematics. The motivating problem is thus to create a foundational system (in the sense of Section 2.1) such that any grammatically well-formed property about a mathematical object is invariant under the appropriate criterion of identity for that object (as those are formalized in the given system).

With this in mind I define a structuralist foundation of mathematics to be a foundational system $\mathcal{S}$ that satisfies the following property:

\[(\text{SFOM}) \text{ Any theoretical context can be naturally formalized in } \mathcal{S} \text{ in such a way that any grammatical property of an object in } \mathcal{S} \text{ is invariant under the relevant criterion of identity in that context.}\]
I will now clarify some of this terminology. A “foundational system” refers to some formal system sufficiently expressive to encode all of (currently-practiced) mathematics, i.e. component (2) of a foundation of mathematics as outlined in Section 2.1. “Object” refers to a mathematical object (e.g. a group, a number, a manifold) and a “grammatical property” refers to a correctly formed expression in $S$ (e.g. a well-formed formula $\phi$ in FOL$_=\). By “theoretical context” I mean a particular mathematical (sub-)discipline (e.g. group theory or topology). By “criterion of identity” I mean the notions of “sameness” employed by practitioners of those mathematical disciplines (e.g. group isomorphism or homeomorphism of spaces).$^{15}$ Finally, “invariance” of a grammatical property $P$ means that if $P$ holds of $x$ in $S$ and if $x \cong y$ in $S$ then $P$ also holds of $y$ (where $x, y$ and $\cong$ are the formalizations in $S$ of two objects in, and the criterion of identity of, a given theoretical context).

It is important to be clear that (SFOM) asserts the existence of invariant formalizations, i.e. those that satisfy the invariance property outlined in its statement. (SFOM) does not assert that all possible formalizations are invariant. As the last sentence makes clear, (SFOM) is a principle relating mathematical practice to the formalization of mathematical practice in $S$. In order to make sense of (SFOM) these two concepts – practice vs. formalization of practice – must be kept apart. For example, the “natural numbers” as understood by the practicing number-theorist are not to be identified with a model of Peano Arithmetic studied by the set theorist. Thus, the “invariance” that (SFOM) demands is a statement about objects and criteria of identity as formalized by $S$. The

$^{15}$It is an interesting question whether a choice of context determines the criterion of identity or vice versa, but I will not get into this here. Furthermore, it is clearly an oversimplification to say that mathematical disciplines are rigidly demarcated by a specific type of structure they study up to a specific criterion of identity. For instance, how would one demarcate number theory in those terms? Or algebraic geometry? Nevertheless, nothing of substance hinges on the accuracy of my sociological analysis – its purpose is merely to clarify the concepts of “theoretical context” and “criterion of identity” as they will be used below.
dominance of set theory in the last half-century has often tempted philosophers to identify mathematical concepts with their set-theoretic formalizations – it is of paramount importance that we resist this temptation here.

Furthermore, we must note that the mere existence of invariant formalizations is too weak a demand to prove interesting. After all, almost any foundational system can reasonably claim the mere existence of invariant formalizations, since anything can be called a formalization, however contrived. For example, we could say that Peano Arithmetic provides invariant formalizations for mathematical contexts simply by coding mathematical statements as (lists of) numbers in a particular way. To avoid such cases, (SFOM) imposes the additional constraint of requiring the existence of “natural” formalizations. This constraint demands that the required invariant formalizations are:

(1) **Uniform**: As much as possible, the invariant formalizations are obtained through some general method of encoding that applies equally well to all mathematical structures of interest

(2) **Native**: The practicing mathematician would find the formalized notions obtained through these invariant formalizations easily comprehensible

One may summarize the requirement of naturality as follows: (SFOM) demands both the natural existence of invariant formalizations (*uniformity*) and the existence of natural invariant formalizations (*nativity*).

Finally, let me note that as a criterion for foundational systems, (SFOM) clearly depends on the current practice of mathematics. For it is mathematical practice that determines not only which notions have to be formalized but also which formalizations are native and whether or not they are invariant. As such, the ambition of (SFOM) cannot exceed our grasp of current mathematical practice. Mathematical practice evolves and so do the structures that mathematicians find useful to work with. So the most
reasonable way to understand (SFOM) is as the demand that for most of the mathematical structures of current interest there exist a uniform method for producing native invariant formalizations. To speak of a foundation that can do this for all mathematical structures of past, present and future interest is a goal worthy only of an incredulous stare – at least insofar as one remains grounded in the practice of mathematics, as we intend to do.

Indeed, (SFOM) is put forward as an ideal to aspire to rather than a realistic goal. In that sense, it is similar to the ideal of more traditional foundations, summarized nicely in [141] as the attempt to find “consistent, finitely axiomatizable theories [into which as many theories can be interpreted].” Full assurance of consistency is never attainable by Gödel’s theorems, but considerations of simplicity, coherence etc. may provide better or worse reasons to believe in it. Analogously, it is hard to imagine that (SFOM) will ever be decisively satisfied, i.e. for any theoretical context. Rather, we will seek foundational systems into which there are natural invariant formalizations of ever more theoretical contexts. What I will establish in this Chapter is that UF achieves just that, i.e. it comes closest to (SFOM) than all other foundational proposals even if it remains shy of the ideal. Nevertheless, UF represents a watershed moment since it appears that if one is to make any progress with respect to (SFOM) one has to move beyond a broadly Cantorian picture of the foundations of mathematics. To see this, we will now examine the extent to which current set-theoretic foundations fail (SFOM).

2.2.1. ZFC and ETCS are not structuralist foundations. It is relatively straightforward to see how ZFC fails (SFOM).\(^\text{16}\) Let our “theoretical context” be traditional group theory, i.e. the study of groups up to isomorphism exemplified by the kind of methods and questions that, say, go into the monumental proof of the classification of finite simple groups. In this theoretical context the “objects” of study are groups and the “criterion

\(^{16}\)Everything I say applies to any standard set-theoretic foundation. I pick ZFC for this example merely for its “brand recognition”.

83
of identity” is group isomorphism. In ZFC, a group is formalized by first-order formulas defining what it is for a set to have group structure. More precisely, we define a three-place predicate

$$\text{Group}(x, *, 0) \equiv * \in (x \times x \rightarrow x) \land 0 \in x \land \text{GroupAxioms}(*, 0)$$

with the obvious abbreviations. What it is to be a group in ZFC is to satisfy this predicate. Similarly, what it is for two groups to be isomorphic is formalized by a predicate

$$\text{Iso}(x, y) \equiv \text{Group}(x) \land \text{Group}(y) \land \exists f \in (x \rightarrow y)(\text{homom}(f) \land \text{bijective}(f))$$

again with the obvious abbreviations and suppressing explicit mention of the group operations.\(^{17}\) What it is for two sets \(x, y\) to be isomorphic as groups in ZFC is to satisfy \(\text{Iso}(x, y)\). Furthermore, by a “grammatical property of an object” in ZFC we mean simply a one-place open formula \(\phi(x)\) in the language of ZFC and for such a grammatical property \(\phi(x)\) to be “invariant under the relevant criterion of identity” we mean that

$$\text{ZFC} \models (\text{Iso}(x, y) \land \phi(x)) \rightarrow \phi(y)$$

So with all this in mind now take

$$\phi(x) \equiv 1 \in x$$

\(^{17}\)I am being a bit quick here, for the sake of exposition. Strictly speaking, the actual operation and identity elements of groups \(x, y\) would have to be used in defining what it is for a function \(f\) to be a homomorphism.
to be such a “grammatical property” and consider \( \mathbb{Z} \) and \( 2\mathbb{Z} \) with their canonical (additive) group structure.\(^{18}\) We have

\[
\text{ZFC} \models \text{Iso}(\mathbb{Z}, 2\mathbb{Z})
\]

But we also have

\[
\text{ZFC} \models \phi(\mathbb{Z})
\]

whereas

\[
\text{ZFC} \nvdash \phi(2\mathbb{Z})
\]

In ZFC therefore we have that \( \phi(\mathbb{Z}) \) is true whereas \( \phi(2\mathbb{Z}) \) is not true even though \( \mathbb{Z} \cong 2\mathbb{Z} \), i.e. even though \( \mathbb{Z} \) and \( 2\mathbb{Z} \) are identical with respect to the relevant group-theoretic criterion of identity. In short, grammatical properties in ZFC are not invariant with respect to group isomorphism as it is formalized in ZFC. Therefore, ZFC fails (SFOM).\(^{19}\)

The above analysis carries over almost verbatim to so-called “structural set theories”\(^{96, 132}\) of which Lawvere’s ETCS\(^{95}\) is perhaps the paradigm. To see this, let \( o \) stand for the canonical arrow

\[
o : 1 \to \mathbb{Z}
\]

and let

\[
\phi(x) \equiv \text{cod}(o) = x
\]

\(^{18}\)By “1” here I mean the set corresponding to the successor of zero in some choice of a model of arithmetic inside ZFC, e.g. the singleton set \( \{\emptyset\} \). By \( \mathbb{Z} \) and \( 2\mathbb{Z} \) I mean the formalizations of the additive groups of integers and even integers respectively.

\(^{19}\)Of course, one could argue that there are non-standard ways of formalizing groups and group isomorphisms in ZFC such that the required invariance holds. However, it is highly doubtful that such methods would be natural (i.e. uniform and native) in the sense demanded by (SFOM).
where \textbf{cod} is the codomain operation of arity \textbf{Arrows} → \textbf{Objects} that takes an arrow to its codomain.\(^{20}\) Then it is immediately seen that

\[ \text{ETCS} \models \phi(Z) \]

whereas

\[ \text{ETCS} \not\models \phi(2\mathbb{Z}) \]

Thus, just like ZFC, ETCS also fails (SFOM).

That it does so illustrates an important omission on the part of those who have advocated ETCS (and its variants) as “structural” set theories. A commonly-employed argument in making such a claim is that in ETCS the natural numbers \( \mathbb{N} \) are only defined \textit{up to isomorphism}. More precisely, in ETCS one defines the natural numbers by defining a natural numbers object (NNO) via a universal property. Thus, one of the axioms of ETCS asserts that such an NNO satisfying the requisite universal property exists. It is indeed the case that this axiom determines an NNO only up to isomorphism – but then again so does the axiom of infinity in ZFC “determine” \( \aleph_0 \) up to bijection. And just as the axiom of infinity in ZFC does not prevent us from expressing properties that are true of one countably infinite set but not of another, so does the axiom for the NNO not prevent us from expressing properties that are true of one NNO but not of another (necessarily) isomorphic to it. Similarly, the axiom of an NNO no more prevents us from stating properties that separate \( \mathbb{Z} \) from \( 2\mathbb{Z} \) than the axiom of infinity prevents us from stating properties separating \( \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \ldots\} \) from \( \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \ldots\} \). This indicates that the real culprit is not the notion of elementhood of extensional set theories like ZFC but rather the availability of a global (untyped) identity predicate. In the case of ZFC

\(^{20}\)I am glossing over the details of the exact presentation of ETCS as a many-sorted first-order theory. For a clear and concise presentation see Palmgren [143].
this identity is intended to apply to individual sets and in the case of ETCS it is intended to apply to individual functions between sets, but this makes little difference with respect to (SFOM).

That said, an advocate of ETCS may point out that the failure of (SFOM) sketched above relies on the non-invariance of an *impure* property, namely one that involves an assertion of identity concerning the object under consideration (e.g. \( \text{cod}(o) \) in \( \phi \) as above). Surely, one can then object, chastising ETCS for its failure to make impure properties invariant is not the same thing as chastising ZFC for its failure to make properties such as \( 1 \in x \) invariant – for after all why not raise the exact same criticism for ZFC by relying on an impure property such as \( Z = x \)?\(^{21}\) Furthermore, it is well-known that one can actually prove that ETCS does satisfy an invariance scheme for all properties \( \phi \) that do not involve the assertion of such impure properties. Therefore, ETCS seems to do much better than ZFC here: the only properties of objects that are not invariant under isomorphism are impure properties – and we can canonically avoid talking about them.

I have three things to say in response. Firstly, if one is not allowed to state in a structural set theory – with an “arrows” ontology – that a certain object is the codomain of a certain arrow, then one really is missing out on properties crucial to one’s ontology. For after all, what defines an arrow in category theory other than its domain and codomain? Secondly, though the distinction between pure and impure properties can be made perfectly precise (one can even inductively define impure properties as a subset of all properties), this does not provide a *reason* to simply ignore impure properties. In other words, these impure properties are a proper part of the universe of sets defined by ETCS – objects in that universe *can* be distinguished on that basis. Thirdly, and I think most importantly, if we allow ourselves such a move of “limiting expressive power” by

\(^{21}\)I owe this objection to Colin McLarty, from whom I have also borrowed the term “pure” and “impure” property.
considering only certain properties but not others (in the case of ETCS pure over im-
pure ones), then why not simply distinguish between properties that are invariant under
isomorphism and those that are not? We can perfectly well do this in ZFC – and there
is therefore no point even to move to a structural set theory like ETCS. In other words,
as soon as we open ourselves up to distinguishing between well-formed formulas based
on some external principle – however precise and motivated it may be – then we might
as well distinguish them on the external principle that is more relevant to our purposes:
whether they are or are not invariant under isomorphism. And this would immediately
make ZFC a structural foundation too – which is another way of saying that we have
allowed ourselves too much.

One might further object that this criticism of ZFC and ETCS relies essentially on
their being formulated in first-order logic. This is correct: the criticism I have presented
so far applies to languages with a global (untyped) identity predicate and not just to
certain theories that can be formulated in terms of such languages. There are indeed
alternative ways of formulating some of these theories that make no essential use of first-
order logic. For example, ETCS can be formulated on top of a simple dependent type
type theory or, alternatively, as a theory in Makkai’s First-Order Logic with Dependent Sort
(FOLDS).

Yet although such formulations do avoid the kind of objection outlined above for
set-level structures (such as \(\mathbb{Z}\)) they still do not avoid it for higher-level structures like
categories. The reason, essentially, is once again that such formulations contain something
that behaves very much like an untyped identity predicate. To illustrate, consider the
axiomatization of ETCS as the FOLDS \(\mathcal{L}_{\text{abset}}\)-theory \(T_{\text{abset}}\) of abstract sets in [113].
One can show that any \(\mathcal{L}_{\text{abset}}\)-sentence is invariant under isomorphism of diagrams in
any model \(\mathcal{E}\) of \(T_{\text{abset}}\). Thus, any set-level structure that can be expressed in terms of
an internal diagram in $\mathcal{E}$ will satisfy the required invariance property. Formalizing such set-level structures in terms of diagrams in $\mathcal{E}$ thus gives us a (SFOM)-compatible general method for formalizing all set-level structures. However, this invariance does not extend beyond isomorphism of diagrams, and therefore becomes problematic when we want to axiomatize higher-level structures like categories.

To see this, recall that a category internal to another category with appropriate structure is given by a diagram

$$
\begin{array}{cccccc}
\vdots & \rightarrow & C_1 \times C_0 & \rightarrow & C_1 & \rightarrow & C_0 \\
& & m & & d & \\
& & & & c & \leftarrow
\end{array}
$$

satisfying certain conditions (where $d$ is for “domain”, $c$ is for “codomain”, $i$ is for “identity” and $m$ is the composition operation). As explained above, FOLDS $\mathcal{L}_{\text{abset}}$-sentences will be invariant only under isomorphism of such diagrams. This means that the desired invariance will hold only if the sets of objects of two such internal categories are isomorphic. Thus, $\mathcal{L}_{\text{abset}}$-formulas will be invariant only up to isomorphism, rather than equivalence, of categories and it is the latter that is usually the criterion of identity one wants to consider in this theoretical context.\(^{22}\) So in particular in $\mathcal{T}_{\text{abset}}$ it is possible to define the category $\mathbf{1}$ consisting of one object and one identity morphism and the category $\mathbf{1}_i$ with two objects and an isomorphism between them, and then to state an $\mathcal{L}_{\text{abset}}$-sentence that is true of one but not true of the other. The same reasoning can be repeated for similar such “type-theoretic” axiomatizations of ETCS, thus casting a doubt on their adherence to (SFOM) beyond set-level structures.

One point that may be granted, however, is that structural set theories like ETCS certainly do better than ZFC in this respect since it is not even possible to formulate the latter in this “half-way” manner. For even if ZFC is formulated on top of a type theory,
the culpable distinctions can still be made from the very beginning because we will have available to us the membership relation “∈”. Nevertheless, neither approach, whether fully first-order or “half-way”, manages to avoid making irrelevant distinctions at some point: they all do eventually manage to distinguish isomorphic objects (for appropriately “high-level” notions of isomorphism).

Another objection is that the kind of argument I am offering against ZFC and ETCS is actually unfaithful to mathematical practice. In other words, one might object that I am chastising ZFC and ETCS on grounds that might seem dubious to the practicing mathematician. For after all, practicing mathematicians are perfectly willing to recognize non-identical isomorphic objects. Indeed no practicing mathematician would regard the even numbers as identical to the natural numbers simply because they are isomorphic as sets. And even in the case of group theory, on which the above argument is based, it is sometimes essential to distinguish isomorphic objects, e.g. when one wants to count the number of subgroups of a certain group. For if one were to regard isomorphic groups as actually identical then one would not be able to say, for example, that the Klein 4-group has two distinct subgroups of order 2 because these two subgroups are isomorphic qua groups.

A cheap way to avoid this objection is simply to say that studying subgroups of groups constitutes a different theoretical context than group theory, in which case the correct notion of identity between the objects of study is not group isomorphism but group isomorphism compatible with certain (distinct) embeddings into some larger ambient group. Formally, this is indeed exactly what prevents one from being trapped in a situation where one is unable to distinguish between isomorphic subgroups (e.g. in systems like ETCS and HoTT). But the cheapness of this response consists in that the vast majority of practicing mathematicians are unlikely to think of subgroups in this manner.
Instead, such mathematicians are far more likely to think of the objects of their study (if these are subgroups) as distinct but isomorphic entities, entities whose isomorphism classes one may wish to *count* (e.g. in order to say that the Klein group has 3 distinct subgroups of order 2, one for each element of order 2). Which raises a crucial point: the more widely we apply the mantra that isomorphic objects should be treated as identical, the farther away we seem to move from the kind of faithfulness to practice that originally motivated this investigation.

That may be so, but there is a related pattern of reasoning that, I submit, really does figure in the thinking of practicing mathematicians. This is the idea that relevant properties can be *transferred* along isomorphisms. Namely, that insofar as an isomorphism has been established between two mathematical structures of interest, everything one may assert of one may also be asserted of the other, as long as such an assertion refers to the *structure* that the established isomorphism is defined in terms of. Thus, I claim, even though a practicing mathematician will find it natural to distinguish between isomorphic subgroups of a certain group (e.g. between \(\langle a \rangle\) and \(\langle b \rangle\) as isomorphic but distinct subgroups of \(V\) where \(a\) and \(b\) are the two distinct elements of \(V\) of order 2), she will also find it natural to distinguish between properties these subgroups enjoy *qua* groups (e.g. that \(\langle a \rangle\) is cyclic) and properties they enjoy *qua* subgroups (e.g. that \(b \notin \langle a \rangle\)). When asked, however, to make this distinction precise, the practicing mathematician will be hard-pressed to find a principle in set theory that allows her to do that – I think it is likely that such a mathematician will at first invoke a certain “obviousness” and if pressed harder may perhaps offer an explanation in terms of properties “expressible in the language of groups” vs. “properties of the ambient set theory”. But however compelling such an explanation may be, it still not a *formal principle* of foundational systems like ETCS.
and ZFC. There are various ways one can prove metatheorems encoding such formal principles (e.g. model-theoretic results about invariance under elementary equivalence) but such results cannot be invoked as fundamental principles of a foundational system, but are rather to be regarded as pieces of mathematics one can encode in terms of them.

So, although I do have to concede that the more seriously\(^{23}\) one takes the naïve mantra of making isomorphic objects indistinguishable the less faithful one is to how practicing mathematicians think of the objects of mathematics, I also maintain that one does regain a measure of faithfulness with respect to how practicing mathematicians think of certain distinctions in the language they use to express properties (and structure) of such mathematical objects. Thus, insofar as we are going to make any progress on (SFOM) while also retaining some faint claim that we are remaining faithful to mathematical practice it appears that we need to entirely reconsider the basic language on which current foundations are based, rather than this or that axiom describing their objects we can express using it.

2.2.2. From Ontology to Language. Clearly, then, (SFOM) is not at all intended to separate set-theoretic foundations from category-theoretic foundations of mathematics. As such, the demand for (SFOM) transcends the debate between category-theoretic and set-theoretic foundations of mathematics. But have we gone too far then? By lumping category-theoretic and set-theoretic foundations together, what kind of circle have I drawn around them? Is there anything left on the other side?

In a way, (SFOM) urges a “linguistic turn” in the foundations of mathematics. Forget about the metaphysical view that a certain foundation takes of its basic objects. Instead, focus on what properties of those objects its language allows it to express. To be sure, there is a significant metaphysical difference in how a material set theory and a structural

---

\(^{23}\)Or “recklessly” to borrow a phrase of John Burgess.
set theory view the objects they are trying to axiomatize. But it is their linguistic promiscuity that (SFOM) rejects, not their intended metaphysics. Structuralism in the sense of (SFOM) is thus best thought of as a constraint on language: we want to stop ourselves from being able to express non-structural properties of objects.

One may now wonder: if this constraint on language is all that is required, then why not just replace objects with isomorphism classes? Both set-theoretic and category-theoretic foundations allow us to do this and doing so certainly would satisfy the criterion that (SFOM) demands of us: for if we formally identify two things, then surely we won’t be able to say something of one that is not true of the other since there will not even be an “other” of which to say anything at all.

The problem is that treating isomorphic objects as actually identical involves essential loss of information. To see this, we must first note that there are two ways of treating isomorphic objects as identical. Firstly, by simply lumping them together in isomorphism classes and ignoring non-identity isomorphisms between them. For example, if I were to replace $\mathbb{Z}$ with its isomorphism class as a group, I am no longer able to ask how many homomorphisms there are, say, between $\mathbb{Z}$ and $2\mathbb{Z}$ – indeed I cannot really talk about $\text{Hom}(\mathbb{Z}, 2\mathbb{Z})$ at all. However, that $\mathbb{Z}$ and $2\mathbb{Z}$ are isomorphic as groups is a non-trivial mathematical fact – not being able to express it at all is clearly a crippling limitation.

Secondly, and less naively, we can treat isomorphic objects as identical by identifying isomorphic objects without forgetting about non-identity isomorphisms between them. This is essentially what happens when one moves from a category $\mathcal{C}$ to a skeleton $sk(\mathcal{C})$: isomorphisms between objects $a, b$ in $\mathcal{C}$ become (non-identity) automorphisms in $sk(\mathcal{C})$. Thus, although we have “fewer” objects than we started with we have just as many arrows and this ensures that no essential information has been lost. Unfortunately, the problem reappears when we then want to talk about further structure on $sk(\mathcal{C})$. For example, even
if we move from the category of sets $\textbf{Set}$ to one of its skeletons $\mathcal{S}$ (thus identifying bijective sets) this does not ensure that isomorphic groups defined on $\mathcal{S}$ are now also equal. To achieve this, we would have to then consider a skeleton of the category $\textbf{Grp}(\mathcal{S})$, and so on. This brings into sharp relief the following crucial distinction: identifying isomorphic objects is not at all the same thing as making isomorphic objects indistinguishable. The former involves loss of information, whereas the latter does not. Doing the former is possible even in ZFC or ETCS but doing the latter is not.

In other words, (SFOM) does not express an appeal to identify isomorphic objects. Quite the contrary: it is of paramount importance, mathematically, to be able to state an isomorphism between two mathematical objects, e.g. between $\mathbb{Z}$ and $2\mathbb{Z}$. And in order to even be able to state such an isomorphism, clearly we do need to view $\mathbb{Z}$ and $2\mathbb{Z}$ as distinct on some level. The reader has perhaps already felt some unease about this point. We are speaking of making isomorphic objects indistinguishable, but if an isomorphism between them is to be construed as a substantive statement, then surely we must distinguish them somehow. And if we can distinguish them somehow, even if they are isomorphic, then what kind of structuralists are we? Even worse, if we buy into Quine’s [150] slogan that there should be no entity without identity, then to accept that our isomorphism relates entities is to pre-suppose that there is some notion of identity (other than isomorphism) that allows us to view them as distinct to begin with. We appear to have fallen into a trap.

At this point I want to say that we have been trapped only by an excessive preoccupation with ontology. After all, what (SFOM) demands is quite simple: as long as we can establish an isomorphism between two objects, then there is no grammatical property of one that does not also hold of the other. This says nothing about what these objects are. That is not to say that we don’t know what these objects are: they are
formalized versions of constructions employed by mathematicians. “And what are these constructions that mathematicians employ?” Enough. It is irrelevant to our purposes. All we require is that whatever these constructions are (Platonic abstractions, mental occurrences, intersubjectively valid concepts etc.) the people that use them – the mathematicians – agree that they have been adequately formalized in the given foundational system we are studying. Is it possible to speak of such objects without committing ourselves to a notion of identity that is finer than any isomorphism that we could establish between them a posteriori? I think it is, at least in the following uncontroversial sense: whatever notion of identity we require in order to distinguish between two presentations of isomorphic objects, it is not necessarily a notion of identity relevant in the mathematical practice from which the particular notion of isomorphism emerges. For example, we have a way of presenting $\mathbb{Z}$ as a group and we also have a way of presenting $2\mathbb{Z}$ as a group. But when we are doing group theory these distinct ways of presenting these isomorphic groups should not automatically commit us to regarding propositions like “$\mathbb{Z} \neq 2\mathbb{Z}$” as relevant to group theory.

Therefore, in order to avoid the Quinean trap, we should make this most minimal of pre-suppositions: that there are distinct ways of presenting objects without committing ourselves to the existence of a significant proposition that distinguishes them. This presupposition allows us to say: if we can establish an isomorphism between these two distinct ways of presenting this object, then we should not be able to state anything in our language that is true of one presentation but not true of the other. And this

---

24 This attitude towards mathematical objects is inspired by Burgess’ permanent parameter structuralism as elaborated in [27]. What I am saying, roughly, is that mathematical objects should be understood as parameters, about whose nature we do not care as long as they behave the way we expect them to. And this allows us to speak of “alternative presentations” of such objects without committing ourselves to a criterion of identity between them.

25 Of course, in the context of number theory a proposition like “$\mathbb{Z} \neq 2\mathbb{Z}$” is highly relevant indeed since we care very much to distinguish odd from even integers, e.g. when we state Goldbach’s conjecture. But in the context of number theory group isomorphism is not, in general, the relevant criterion of identity.
is exactly what (SFOM) demands. Even the word “object” in the previous sentence is something of a distraction. We can drop talk of objects altogether. Forget about referents/extensions. All we have are senses/intensions – and what (SFOM) wants is that if two such senses/intensions can be shown to be “isomorphic” then the required invariance property holds. In shifting our focus from ontology to language, we can thus completely invert the Quinean motto: there is no identity without entities. In other words, if we think of “identity” between two objects $A$ and $B$ as no more than what allows us to transfer any properties of $A$ to $B$ and vice versa, then what (SFOM) demands is that “identity” become identified with “isomorphism”. Or, more accurately: that “identity” become *isomorphic* with “isomorphism”. As I will now go on to argue, this is exactly what the axiom of univalence in UF achieves.

### 2.3. Univalence and (SFOM)

What does the axiom of univalence say? It says, quite literally, that to ask of two types whether or not they are identical is the same as asking whether or not they are “isomorphic.” In symbols this can be expressed as follows. For any types $A$ and $B$ (understood as terms in a universe of types $U$) the following holds:

$$(A =_U B) \cong (A \cong B)$$

where “$\cong$” is a symbol for (the syntactic definition of a general notion of) isomorphism and “$=_U$” is the symbol for the identity type in MLTT formed in $U$ (when $A$ and $B$ are regarded as its terms). Thus, when properly paraphrased, the axiom of univalence (UA) asserts the following: *Identity is isomorphic to isomorphism.*

---

26 More accurately, but less sonorously: there is no *criterion of* identity without *names for* entities.

27 As has already been noted, the precise statement of univalence asserts that identity is isomorphic to isomorphism by asserting that a canonical map from identities to isomorphisms is an isomorphism. In other words, univalence asserts that a particular (canonical) map is an isomorphism, and not merely the existence of an isomorphism.
The statement of univalence might immediately appear baffling. For it seems as though UA is asserting an isomorphism between two things, one of which is not at all clearly a structure. For suppose one accepts that the type “\( A \cong B \)" of isomorphisms is a structure of some sort. Then what about the identity \( A =_U B \)? Identity, surely, is a proposition of some sort. It is either true or false. How can a proposition be isomorphic to a structure? It appears that the statement of univalence involves a category mistake.

Strange as it may initially appear, the answer is that in UF (and in intensional MLTT more generally) identity is not (generally) a proposition – it just is a structure. Depending on \( A \) and \( B \), \( A =_U B \) could turn out to be a structure containing a lot more information than merely whether it is inhabited (“true”) or not (“false”). For example, there could be distinct terms of type \( A =_U B \), which one can think of as distinct proofs of the identity between \( A \) and \( B \). In short, for the statement of UA to make sense one must accept the point of view of identity-as-structure. And after doing so, it is not too much of a stretch to view UA as analogous to the axiom of extensionality in set theory. Just as the latter provides a definition of (the previously underdetermined) equality of sets (“Two sets are equal if and only if they have the same elements”) so does UA provide a definition of the (previously underdetermined) identity of types in intensional MLTT. The key difference, however, is that the “definition” that UA provides does not assert the logical equivalence of one fact (“being equal sets”) to another fact (“having the same elements”). Instead, it asserts the isomorphism of one structure (“identity type”) to another structure (“type of isomorphisms”).

So now let us ask: what can univalence do for us when it comes to (SFOM)? Firstly, UA ensures that any grammatical property \( P \) expressible in HoTT that is true of a type \( A \) is also true of a type \( B \) that is isomorphic to \( A \). Following Awodey [9], we may call this
fact the *Principle of Isomorphism* (PI).\(^{28}\) As shown by Awodey \([9]\) in quite some detail, UF satisfies (PI). How? Roughly, because a proof of isomorphism between \(A\) and \(B\) can give rise, through UA, to a proof of identity between \(A\) and \(B\). And since properties are invariant under identity we get that if \(P\) holds of \(A\) then it also holds of \(B\). Thus, any grammatically well-formed property in HoTT is invariant under isomorphism.

Furthermore, we must now note that the term “isomorphism” in the statement of (PI) and in how we have been using it above actually refers to a much more general notion. This more general notion is *homotopy equivalence* which can be understood as the “correct” criterion of identity for homotopy types. The notion of homotopy equivalence originates in algebraic topology but it can be expressed syntactically in intensional MLTT without the addition of extra non-logical symbols. That this can be done is one of the key insights that led to the homotopy interpretation of MLTT. Thus, given any two types \(A, B\) in \(\mathcal{U}\) it is possible to define a type of homotopy equivalences between \(A\) and \(B\) using only the symbols of MLTT. One can certainly continue to think of it as the type of “isomorphisms” between \(A\) and \(B\) as long as it is clear that it is a more general notion than what is usually understood by the term “isomorphism”, i.e. structure-preserving bijections between sets. But – as Awodey himself notes – homotopy equivalence does specialize to more familiar notions in more familiar cases. For example, bijections between sets can be seen as special instances of homotopy equivalence and so can logical equivalence of propositions.\(^{29}\)

An important issue here is that homotopy equivalence, as defined in HoTT, only meaningfully applies to pairs of types \(A, B\) in a (univalent) universe \(\mathcal{U}\) and not to terms of such types. Therefore, it makes *prima facie* no sense to speak of homotopy equivalence

---

\(^{28}\)In \([9]\), Awodey refers to (PI) as the Principle of *Invariance* but this is merely a terminological difference.

\(^{29}\)In keeping with the notation above, I will continue to use \(A \simeq B\) to denote the type of homotopy equivalences, except when ambiguity might arise, as in Section 5, where I will use the more standard “\(\sim\)”.
between two terms $a, b$ of a type $A$ in $\mathcal{U}$. We shall however stipulate here that we will regard such instances as degenerate cases of the notion of homotopy equivalence. In other words, we will allow ourselves to refer to the identity type $a =_A b$ for terms $a, b$: $A$ as the type of homotopy equivalences between them; from now on the term “homotopy equivalence” will be understood in this expanded sense and we will use the notation $a \simeq b$ to denote it. One can of course take this as no more than a terminological convention. But since the way in which I will argue for (CI) below depends on it I ought to say a few words about why it is a reasonable point of view to adopt. The reason, quite simply, is that types themselves can be regarded as terms of a (univalent) universe $\mathcal{U}$ and in that case homotopy equivalence does apply to terms of a type. From a topological point of view, anything can be regarded as a point or as a space; viewing spaces as points in some larger space allows us to talk about homotopy equivalence between points.\footnote{This can even be formalized in point-set topology for sufficiently nice spaces, by regarding their points as singleton subspaces and showing that equal such points are (trivially) homotopy-equivalent as singleton subspaces.}

With this in mind, one cannot fail to notice the similarity between (PI) and (SFOM) when the latter is specialized to UF. For what we have is

(PI) Any grammatical property of an object in HoTT is invariant under homotopy equivalence

and what we need is

(SFOM-UF) Any theoretical context can be naturally formalized in HoTT in such a way that any grammatical property of an object in HoTT is invariant under the relevant criterion of identity in that context
In order to argue that UF is a structuralist foundation in my sense it suffices to show that \( (PI) \Rightarrow (SFOM-UF) \) and as the above comparison makes clear, to do so one needs to argue that

\[(F) \text{ any theoretical context can be naturally formalized in HoTT}\]

and

\[(CI) \text{ in such a way that the relevant criterion of identity for that context coincides with homotopy equivalence.}\]

I will investigate the extent to which \((F)\) is true in Section 2.5. The following Section is devoted to showing that \((CI)\) is true.

2.4. Criteria of Identity as Homotopy Equivalence

In order to establish \((CI)\) we must provide a general method for formalizing mathematical structures such that their formalized criterion of identity coincides with homotopy equivalence. In what sense of “coincide”? At the very least, in the sense of “being homotopy equivalent to” it.

The set-up is as follows: we are given informally a class \(O_{\inf}\) of objects and a specification of a criterion of identity \(\sim_{\inf}\) for these objects.\(^{31}\) A formalization of the theoretical context given by \(O_{\inf}\) and \(\sim_{\inf}\) then consists of a type \(O_{UF}\) in \(U\) together with a relation \(\sim_{UF}\) between any two terms of that type. Formally, we can write this relation as

\[\sim_{UF}: O_{UF} \rightarrow O_{UF} \rightarrow U\]

\(^{31}\)What this “class” is and how exactly it is “specified” we leave open. At the very least, there should be some consensus among practicing mathematicians that studying objects of type \(O_{\inf}\) under \(\sim_{\inf}\) constitutes a legitimate specialization. For instance, studying groups under group isomorphisms (“group theory”) or studying topological spaces up to homotopy equivalence (“homotopy theory”).
and it is important to note that this relation is not necessarily a proposition. Namely, for any \(a, b\): \(O_{\text{UF}}\) the most we can tell about \(a \sim_{\text{UF}} b\) is that it is a type in \(U\).\(^{32}\)

What remains now is to show that \(\sim_{\text{UF}}\) coincides with homotopy equivalence. Clearly there are many ways to make this precise, as we have not restricted ourselves to a single way of understanding what it means for \(\sim_{\text{UF}}\) to “coincide with homotopy equivalence.” I will make it precise in the following way:

i. Define \(\sim_{\text{UF}}\) as a type \(\cong_O\) canonically definable from \(O_{\text{UF}}\) and canonically equivalent to the type of homotopy equivalences on \(O_{\text{UF}}\).\(^{33}\)

ii. Provide a general method that exhibits \(\cong_O\) as the correct formalization of \(\sim_{\inf}\) under a certain way of understanding \(O_{\inf}\) and \(\sim_{\inf}\).

The first task is taken up in Section 2.4.1 and the second task in Section 2.4.2.

2.4.1. Component-wise isomorphism as a criterion of identity. To define \(\cong_O\) we must consider two cases. Firstly, that in which \(O_{\inf}\) is formalized as a type that does not involve collecting other types together. For example, if \(O_{\inf}\) is the class of natural numbers \(\mathbb{N}\) and \(\sim_{\inf}\) is the relation of equality of numbers – in which case the “theoretical context” is, say, number theory. Secondly, that in which \(O_{\inf}\) is formalized as a collection of types with certain additional structure, satisfying certain properties. For example, if \(O_{\inf}\) is the class of groups and \(\sim_{\inf}\) is the relation of group isomorphism – in which case the “theoretical context” is, say, group theory. In model-theoretic terms we can think of types of the first kind as theories with an intended model (e.g. arithmetic) and types of the second kind as theories without an intended model (e.g. group theory).

\(^{32}\)Just like what was said about identity types in the beginning of Section 2.3, a criterion of identity is also to be understood as a structure rather than a proposition. Of course, if \(\sim_{\text{UF}}\) is to be meaningfully regarded as a criterion of identity then we will usually assume that it is at least reflexive, symmetric and transitive (in the type-theoretic sense). But we also allow for criteria of identity that satisfy much stronger properties, e.g. identity systems in the sense of [171], Definition 5.8.3.

\(^{33}\)Where “homotopy equivalences” are understood in the expanded sense outlined in Section 2.3.
In the first case, \( O_{UF} \) will be some type constructed in UF without using a \( \Sigma \)-constructor over \( \mathcal{U} \). In other words, given terms \( a, b : O_{UF} \) their identity type \( a =_{O_{UF}} b \) is not equivalent to some type that itself contains an identity type between two types in \( \mathcal{U} \). To illustrate: if \( O_{UF} \equiv \mathbb{N} \) then the identity type between any two terms \( n, m : \mathbb{N} \) is simply the identity type formed in the usual way. It is not equivalent to some other type that is constructed out of other identity types. Thus, as long as we are formalizing a theoretical context of the first kind, the relevant criterion of identity will be given by the identity type for terms of that type, just as in intensional MLTT. Since we have stipulated that in such cases we will consider the identity type as a degenerate version of homotopy equivalence, we get a formalization of the kind demanded by (CI).

In the second case, we are studying a class of structures best formalized as a collection of types. Here \( O_{UF} \) will be of this general form:

\[
S \equiv \sum_{A : \mathcal{U}} \text{Struc}(A) \times \text{Prop}(A)
\]

i.e. a collection of types \( A \) in a universe \( \mathcal{U} \) that have a certain structure \( \text{Struc}(A) \) satisfying certain properties \( \text{Prop}(A) \).

How can we describe the identity type for two terms \( M, N \) of type \( S \)? Roughly, as follows: the identity type of two terms of a \( \Sigma \)-type is given by the \( \Sigma \)-type of identities between those two terms. In more detail, given two terms

\[
\langle a, p \rangle, \langle b, q \rangle : \sum_x B(x)
\]

we have that

\[
(\langle a, p \rangle = \langle b, q \rangle) \cong \sum_{\pi : a =_A b} (\pi_* (p) =_{B(p)} q)
\]

\(^{34} S \) is of course itself a type, possibly in a higher universe of types.
In other words, identities between terms $\langle a, p \rangle$ and $\langle b, q \rangle$ are equivalent to identities between each of their components. But in order for this statement to make sense grammatically (i.e. be “well-typed”) we need to transfer the information appropriately along the previously given identities, which is what the symbol $\pi_*$ indicates. Univalence then licenses us to replace “$=$” with “$\simeq$” in the formula above. Therefore, the identity type of $O_{UF}$ will coincide with the type of homotopy equivalences between each component. For any given $O_{UF}$ we will denote this type of “component-wise isomorphisms” by $\cong_O$.

As an illustration of how $\cong_O$ is defined, consider the case of graphs. Pre-formally, a graph $G = \langle V, E \rangle$ is a collection of vertices $V$ together with, for any two nodes, a collection of (possibly multiple) edges $E$ between any two vertices $v_1$ and $v_2$. We consider two such graphs $G_1, G_2$ equivalent if there is an isomorphism between $V_1$ and $V_2$ that preserves the number of edges. In UF, we can formalize graphs as the following type

$$O_{\text{graph}} \equiv \sum_{V : U} V \to V \to \text{Set}_{U}$$

Now, take two terms of this type, $\langle V_1, E_1 \rangle$ and $\langle V_2, E_2 \rangle$. We have:

$$\langle \langle V_1, E_1 \rangle = \langle V_2, E_2 \rangle \rangle \cong_{\text{pre-formal}} \sum_{p : V_1 \simeq V_2} p_*(E_1) = E_2$$

As explained above, we now define $\cong_{O_{\text{graph}}}$ by replacing the identity signs on the right-hand side above by the types of of homotopy equivalences, denoted here by “$\simeq$”. So in the example this gives us

$$\langle \langle V_1, E_1 \rangle \cong_{O_{\text{graph}}} \langle V_2, E_2 \rangle \rangle \equiv_{\text{pre-formal}} \sum_{p : V_1 \simeq V_2} p_*(E_1) \simeq E_2$$
As is hopefully clear, $\simeq_{\text{O}_{\text{graph}}}$ is defined canonically from $O_{\text{graph}}$, namely by the two-step process of writing out the identity type and then replacing “$=$” with “$\simeq$”.

It might now appear that the relevant question to ask is whether $\simeq_{\text{O}}$ captures the meaning of our original informal notion $\sim_{\text{inf}}$, whatever that may be. In one sense, this question is impossible to settle. There will always be room for a persistent skeptic to doubt that this has been achieved: “Has the meaning of group isomorphism as mathematicians understand it really been captured by the appropriate instance of homotopy equivalence in UF?” But as Quine long ago made vivid, this kind of skepticism is an artefact of the very process of translating (or, in our case, the process of formalizing) rather than of the particular language we are translating into (in our case, HoTT). And in another sense, the question has a trivial answer. After all, the identity types in UF only “see” those features of the terms being considered that were used to define these terms to begin with. And if we agree that the terms of $O_{\text{UF}}$ adequately capture the features of objects of the class $O_{\text{inf}}$ that we are interested in (e.g. that they are sets, that they have a multiplicative operation etc.) then the identity type will “see” all those features and thus preserve them. And therefore, by univalence, so will $\simeq_{\text{O}}$.

Therefore, the issue is not really about comparing the meaning of $\simeq_{\text{O}}$ and $\sim_{\text{inf}}$. The right question to ask, rather, is the following: Is $\simeq_{\text{O}}$ always deduced from $\sim_{\text{inf}}$ in a canonical way? In cases of set-level structures (e.g. groups, rings) formalized as $h$-sets-with-structure the “Structure Identity Principle” (cf. [171], Chapter 9.8 and especially Definition 9.8.4) already provides a standard way of deriving their criteria of identity

\footnote{Occasionally (e.g. in the case of (pre)categorias, see below) we might be interested in more explicit descriptions of the right hand side. For example, in the case of graphs, using function extensionality and the fact that the $E_i$ are set-valued functions, we can obtain

$$((V_1,E_1) \simeq_{\text{O}_{\text{graph}}} (V_2,E_2)) \simeq \sum_{p: V_1 \simeq V_2} \left( \prod_{x,y: V_1} E_1(x,y) \simeq E_2(p_*(x),p_*(y)) \right)$$

which more closely resembles what we have come to expect graph-isomorphism to mean in practice, namely an isomorphism of the vertex-set that induces isomorphisms on the corresponding edge-sets. Nevertheless, even if the process of “rewriting conveniently” is not canonical, the process of obtaining $\simeq_{\text{O}}$ is, in fact, canonical.}
from their definition, in much the same way that elementary equivalence of models in first-order logic can be derived from the signature of the corresponding first-order theory. And in those cases, the given criterion of identity does in fact coincide with ≅_O. But is there such a general method even for more complicated – and not necessarily “set-level” – structures? This is the task I now take up.

2.4.2. The general method for (CI). To illustrate the task at hand, as well as sketch my proposed method, let us consider the case of category theory. First, let me quickly review how category theory is formalized in UF. A widely accepted definition of a category in UF is the following:

**Definition 2.4.1 ([2,171]).** A univalent category C consists of the following data:

1. A type C: U (“objects”)
2. A dependent type Hom_C: C → C → Set U (“Hom-sets”)
3. A term 1: Π a: C Hom_C(a, a) (“identity”)
4. A term °: Π a,b,c: C Hom_C(a, b) → Hom_C(b, c) → Hom_C(a, c) (“composition”)
5. A term
   \[
   \text{assoc}: \Pi_{a,b,c,d: C} f: \Pi_{g: \text{Hom}_C(a,b)} h: \text{Hom}_C(b,c) \quad h \circ (g \circ f) = (h \circ g) \circ f
   \]
   which witnesses (strict) associativity.
6. A term
   \[
   \text{ident}: \Pi_{a,b: C} f: \Pi_{b: \text{Hom}_C(a,b)} (f \circ 1_a = f) \times (1_b \circ f = f)
   \]
   which witnesses right and left cancellability of identity maps.
(7) The canonical map $\text{idtoiso}_{a,b}: a = b \rightarrow a \cong b$ is an equivalence for all $a, b$, i.e. there exists a term

$$\text{cat}: \Pi_{a,b:C} \text{isequiv}(\text{idtoiso}_{a,b})$$

The data in (1)-(6) express in type-theoretic notation exactly the standard defining structure and properties of a category, i.e. that it consists of a collection of objects (1), a collection of arrows (2), an identity arrow on each object (3) and a composition operation (4) that is associative (5) and for which the identity arrows are left and right inverses (6). Condition (7), on the other hand, is new in UF. It expresses, as we shall see, the requirement that isomorphism between objects inside a category be equivalent to identity between those objects. We will refer to it as the saturation condition.

On the other hand, there is another, more “naïve”, notion that could serve as a formalization of category theory in UF:

**Definition 2.4.2.** A precategory $\mathcal{C}$ consists of data (1)-(6) in Definition 2.4.1.

Precategories and (univalent) categories are not equivalent structures or theories. So it appears that there are two distinct plausible formalizations of category theory in UF. Each of them comes with its own criterion of identity and prima facie each of these criteria of identity are plausible formalizations for categorical equivalence as it is understood in mathematical practice. So we now appear to be in a bind: if there are two inequivalent, 36Although a similar kind of condition was considered already by Hofmann and Streicher in their groupoid model for MLTT, cf. [68].

37But there is also a strong relation between them: roughly, every precategory gives rise to a univalent category (called its Rezk completion in [171]) that is “weakly equivalent” to it. For the categorically-minded reader: the obvious forgetful “functor” from univalent categories to precategories has a “left adjoint”. Perhaps someone may wish to claim that this relation between the two notions shows that there is a kind of equivalence between them. This would challenge the points I go on to make below. But this is the wrong conclusion to draw. Similar relations (“forgetful-free adjunctions”) are borne by classes of structures that on no reasonable account of “equivalence” should we wish to call equivalent. For example, groups and (bare) sets bear this relation.
but equally plausible, formalizations of category theory in UF then how could $\sim_O$ be deduced in a canonical way from $O_{UF}$, given that we cannot even pick the correct “$O_{UF}$”? 

What I want to say now is that we are in a bind only insofar as we assume that both univalent categories and precategories are distinct formalizations of the same informal mathematical notion. It is this dubious assumption which we must now lay to rest. Precategories and univalent categories, I submit, are best understood as (distinct) formalizations of distinct notions. Strange as this may seem from a set-theoretic point of view, it is in fact consistent with one of the fundamental tenets of UF, namely that (higher) categories are structures on (higher) groupoids rather than groupoids being categories with an extra property (all arrows are invertible). From this point of view, it is entirely natural that there should be “two (or possibly more) category theories” since a category is no longer a fundamental thing that we can only define in one way, but rather one of many possible structures that we can impose on (higher) groupoids. By analogy, consider the many kinds of orderings we have on bare sets: partial orderings, total orderings, well-orderings etc. Each of these can be studied in set theory as a separate notion. They formalize different kinds of orderings that we may consider in mathematical practice. Would we want to criticize set theory for being unable to capture the “theory of orders”? No, because in set-theoretic foundations sets are fundamental and ordered sets are non-fundamental structures we define on sets. Similarly, in univalent foundations, homotopy types are fundamental and categories are non-fundamental structures we define on homotopy types.\(^{38}\)

\(^{38}\)Accepting the distinction between categories up to isomorphism and categories up to equivalence depends on some level on accepting that groupoids are more fundamental than categories. I argue that this is indeed the case on purely philosophical grounds in Chapter 4 and I will therefore say no more about it here.
What is now left to show is that $\simeq_{\text{catiso}}$ and $\simeq_{\text{catequiv}}$ are derived from $\sim_{\text{catiso}}$ and $\sim_{\text{catequiv}}$ in a canonical way (that applies equally well to both). To see this, we must somehow describe categories-up-to-equivalence and categories-up-to-isomorphism as made up of components in such a way that both $\sim_{\text{catiso}}$ and $\sim_{\text{catequiv}}$ can be understood as asserting a component-wise isomorphism. In the case of categories-up-to-isomorphism, this is immediate since an isomorphism of categories is defined as a functor that is an isomorphism on both objects and arrows. In the case of categories-up-to-equivalence the situation is more difficult. Here an equivalence of categories is usually defined as a functor that is an isomorphism on arrows but only an isomorphism-up-to-isomorphism (i.e. only essentially surjective) on objects. But such a functor, I submit, can also be understood as an abbreviated way of asking for a component-wise isomorphism if we take seriously the fact that we only care about objects in a category-up-to-equivalence up to isomorphism. In other words, the object part of a functor witnessing an equivalence between categories should be understood as providing an isomorphism of objects-up-to-isomorphism, rather than an isomorphism-up-to-isomorphism of objects. And if we see it this way, then the informal criterion of identity $\sim_{\text{catequiv}}$ for categories-up-to-equivalence is of exactly the same kind as the criterion of identity $\sim_{\text{catiso}}$ for categories-up-to-isomorphism.

As a result, if we formalize $O_{\text{catiso}}$ as the type of precategories and $O_{\text{catequiv}}$ as the type of univalent categories, then we can argue that $\simeq_{O_{\text{catiso}}}$ and $\simeq_{O_{\text{catequiv}}}$ are canonically obtained from $\sim_{\text{catiso}}$ and $\sim_{\text{catequiv}}$ for the exact same reasons. Indeed this point of view correlates very well with the formalization of category theory in [171]: isomorphism of precategories (Definition 9.4.8) is easily seen to be equivalent to $\simeq_{O_{\text{catiso}}}$ (Lemma 9.4.14, essentially) and the definition can be repeated in the case of univalent categories. Now it so happens that because of the extra saturation condition, $\simeq_{O_{\text{catequiv}}}$ in the case of univalent categories is also itself equivalent to a relation that more closely resembles
what usually goes under the name of “equivalence of categories” in mathematical practice (Lemma 9.4.14). But that is beside the point here. What matters, rather, is that both isomorphism of precategories and “equivalence” of univalent categories can be obtained, up to equivalence, as component-wise isomorphisms in exactly the same way.\textsuperscript{39}

The above discussion illustrates very well the general method of formalizing criteria of identity as \textit{component-wise isomorphisms} where, as ever, we understand “isomorphism” as a placeholder for the notion of homotopy equivalence. It is now time we stated this general method. For a given theoretical context given by \( O_{\text{inf}} \) and \( \sim_{\text{inf}} \) it goes as follows:

\begin{enumerate}
\item \textit{Decomposition}: break down the structures of \( O_{\text{inf}} \) into all their relevant components \((S_1, ..., S_n)\), some of which may depend on others.
\item \textit{Component-wise isomorphism}: determine the relevant “local” criteria of identity \((\sim_1, ..., \sim_n)\) for objects in each of these components in such a way that the “global” criterion of identity \( \sim_{\text{inf}} \) can be understood as a component-wise isomorphism.
\item \textit{Saturation}: formalize \( O_{\text{inf}} \) in such a way that the formalized “local” criteria of identity \( \sim_i \) for each formalized component \( S_i \) are homotopy equivalent to the identity type.
\item \textit{Invariance}: Argue that \( \cong_O \) is the correct formalization of \( \sim_{\text{inf}} \).
\end{enumerate}

To illustrate how this general method works it is helpful to sketch how when applied to the theoretical context of categories-up-to-equivalence it gives us the formalized notion of univalent categories in HoTT:

\textsuperscript{39}The notion of a \textit{strict} category ([171], Definition 9.6.1) which is a precategory where the type of objects is an \( h \)-set is perhaps a better formalization of the informal notion of category-up-to-isomorphism. So it is worth noting that everything I have said so far could be repeated for strict categories instead of precategories, changing nothing of substance.
(1) $O_{\text{inf}}$ is given by structures with the following components: Objects ($Ob$), Arrows ($A$), Identities ($I$) and Composition ($\circ$). These satisfy the usual properties that define a category.

(2) The criterion of identity $\sim_{Ob}$ for objects is isomorphism defined in the usual way in terms of $A$, $I$ and $\circ$. The criteria of identity for all the rest of the components of the structure are standard. Categorical equivalence can then be understood as a component-wise isomorphism as sketched above.

(3) We formalize $O_{\text{inf}}$ as the type of univalent categories $O_{\text{catequiv}}$, where condition (7) provides the required saturation condition between the formal version of $\sim_{Ob}$ and the identity type of $Ob$.

(4) As argued above, $\simeq_{O_{\text{catequiv}}}$ is then the correct formalization of $\sim_{\text{inf}}$.

Needless to say, the above-described general method is not a formal algorithm. It is meant, rather, as a general rule of thumb about how to produce invariant formalizations of theoretical contexts in UF. As a rule of thumb, it will require specific tweaks in specific cases and of course will be subject to varying interpretations. I will make no claim here that this method is precise enough to pin down a unique formalization of every theoretical context. But it is, I submit, precise enough to establish a general approach to invariant formalizations – and no more than that is required as an argument in support of (CI).

With this qualification in mind, we must now ask the much more difficult question: does the method for determining criteria of identity as component-wise isomorphisms extend to most structures of interest in mathematical practice? Moreover, does it produce formalizations that mathematicians would find native? To this, and more, the next section is devoted.
2.5. Formalizing all of mathematics in HoTT

To satisfy criterion (F), HoTT must at the very least prove capable of encoding all of mathematics as it is currently being practiced, i.e. it needs to be a “big-f” Foundation for all of mathematics rather than a “small-f” foundation for a specific mathematical discipline. Therefore, I will begin my argument for (F) by countering the three main objections that have been put forward against UF’s capacity as a “big-f” Foundation. Then I will argue that the general method outlined in Section 2.4 can be used to formalize most theoretical contexts and that, in addition, the way in which this is achieved would not appear alien to practicing mathematicians.

2.5.1. Is HoTT a “big-f” Foundation for all of mathematics? As a formalization of the Univalent Foundations, HoTT is based on an extension of a well-understood formal system (intensional MLTT) that itself has been plausibly discussed as offering a self-sufficient foundation for all of mathematics. As such, the question of whether or not HoTT can be regarded as a “big-f” Foundation for all of mathematics seems to me – if not unfair – then at least to be one in which the burden of proof lies firmly on the side of the skeptic. Therefore, I will not provide a positive account for an affirmative answer to the title question, but content myself with countering what I take to be the most widespread objections to such an affirmative answer.

(A) “HoTT cannot be a foundation for all of mathematics because not all of mathematics is homotopy theory!”

Response to (A): This objection (the most common one) betrays a basic misunderstanding about the distinction between the analytic and the synthetic aspect of UF. UF certainly contains a synthetic theory of ∞-groupoids40 and such a theory is certainly

---

40For a precise and thorough explanation of the way in which HoTT/UF is a synthetic theory of ∞-groupoids see Shulman [165].
about (abstract) homotopy theory. But UF also contains an “analytic” (for lack of a better term) version of set theory, namely that given by $h$-sets (i.e. 0-types). Therefore, the short and sweet response to (A) is simply this: UF is homotopy theory only insofar as it constitutes a synthetic theory of $\infty$-groupoids. But UF is not merely a synthetic theory of $\infty$-groupoids. And it is exactly all its other features (including its “analytic” set theory) that make it a viable candidate to serve as a (structuralist) Foundation of mathematics. To put the point differently: UF is not merely a theory about homotopy types; it is a theory allowing you to carry out constructions on homotopy types.\footnote{It is important to note that among such constructions are also those of homotopy types that can be understood as models of set theory as traditionally conceived. For example, the type $\text{Set}_U$ of $h$-sets without any further assumptions on HoTT is a model of a weak predicative set theory (a “$\Pi W$-pretopos”) and if we assume LEM then a model of ZFC can also be constructed as a higher inductive type (cf. \cite{171}, 10.5).}

Yet even though (A) fails as an objection, it does still highlight an important issue: the synthetic and the “analytic” side of UF (formalized as HoTT) do not necessarily interact as one would expect. Let me illustrate with an example. Suppose I am interested in proving that the topological circle $S^1$ is not homeomorphic to the topological sphere $S^2$. I then define $S^1$ and $S^2$ synthetically as higher inductive types (call them “synthetic $S^1$” and “synthetic $S^2$”) and then calculate their fundamental groups in the now well-known way outlined in \cite{171}. Have I thereby proved that topological $S^1$ is not homeomorphic to topological $S^2$? Inside HoTT, I have not.\footnote{This is because it does not even make sense to consider synthetic $S^1$ and $S^2$ “up to homeomorphism”. The very statement “$S^1$ is not homeomorphic to $S^2$” cannot be stated in UF-regarded-as-a-synthetic-theory-of-$\infty$-groupoids. It must be stated with $S^1$ and $S^2$ defined topologically, i.e. as $h$-sets with appropriate extra properties and structure mirroring the way in which these spaces are usually defined in point-set topology.} Which leaves the obvious worry: if in order to draw conclusions about spaces up to homeomorphism you then have to repeat the relevant proofs (e.g. of $\pi_1(S^1) = \mathbb{Z}$) at the level of $h$-sets, then how much is the synthetic side of UF adding to UF-regarded-as-a-foundation?

I think this is an important limitation of HoTT but no deal-breaker with respect to UF’s capacity to serve as a foundation for mathematics, for two reasons. Firstly, the
fact that synthetic calculations of homotopy groups in HoTT cannot be used “as is” to formalize traditional inferences we might draw about spaces up to homeomorphism does not mean that these synthetic calculations are devoid of mathematical merit. Nor does it mean that the above-described inferences (from the non-isomorphism of fundamental groups to the non-homeomorphism of spaces) are not formalizable at all: they can still very well be formalized at the level of h-sets (by repeating the classical proofs). Secondly, it is worth remarking that there could be possible extensions of HoTT that could ameliorate this situation. Most prominently, there is a version of HoTT called cohesive HoTT (cf. [158,159,164]) that makes the interaction between the synthetic and the “analytic” side more robust, essentially by adding a kind of “topological” (or “cohesive”) structure to bare homotopy types. This extra structure allows one to recover the “analytic” versions of mathematical structures from their synthetic versions. In particular, one can prove (cf. [164]) that the realization of the synthetic $S^1$ is equivalent to the topological circle, thus solving the problem described above. Perhaps this means that an argument could be made that cohesive HoTT – even though motivated by questions in the foundations of higher gauge theories in physics – could actually provide a worthy alternative “foundational” formalization of UF. At the very least this demonstrates that the lack of interaction between the analytic and the synthetic side in the current formulation of HoTT is not an insurmountable obstruction.

(B) “UF cannot be a Foundation for mathematics because its basic objects are not fundamental. $\infty$-groupoids were initially defined analytically within set theory as sets with extra structure and properties. Therefore, sets are more fundamental than $\infty$-groupoids.”

Response to (B): The objection here is that a synthetic theory of a certain type of structure (e.g. $\infty$-groupoids in the case of UF) cannot be a Foundation for the whole
of mathematics if that type of structure can be defined analytically in another formal system (e.g. \(\infty\)-groupoids as Kan simplicial sets in ZFC). But this argument can just as easily be turned on its head: by the same token one can argue that since UF provides an analytic description of a set (as an \(h\)-set) then set theory cannot be a Foundation since its basic objects can be defined analytically in another formal system (as \(h\)-sets in UF).

The debate will then come down to which type of structure is more “intuitively” fundamental: “shapes” or “collections”. But this no longer has anything to do with (B). Furthermore, being able to define \(X\) in terms of \(Y\) seems to me a highly unreliable guide to \(X\) or \(Y\)’s relative fundamentality: we can define lines in terms of points (e.g. in Cartesian geometry) but we can also define points in terms of lines (e.g. as the intersections of lines) and these facts do not seem at all relevant in deciding whether points or lines are more fundamental. As for historical reasons of priority (“sets are more fundamental than \(\infty\)-groupoids because they were the first to be thought of as the basic objects of mathematics”) I can only say that the correlation of historical precedence and fundamentality is tenuous at best: would we similarly want to say that point-shaped particles are surely more fundamental than strings because point-shaped particles were the first, historically, to be thought of as the fundamental constituents of matter?

Nevertheless, beyond relative fundamentality, one can also question the inherent simplicity of homotopy types. In other words, are homotopy types objectively simple enough to serve as the basic entities of a foundation of mathematics? I believe the way to answer this is to ask the same question of the criterion of identity which defines homotopy types, viz. homotopy equivalence. For the simplicity of a certain concept is surely positively correlated to the simplicity of its identity conditions. Now a homotopy equivalence is simply a continuous (“smooth”, though not in the sense of calculus) deformation of one space into another. To formalize this idea takes a lot of effort, especially in set theory,
but this does not detract from its conceptual simplicity. And since homotopy types are supposed to be exactly what gets preserved by homotopy equivalence, they share in the simplicity of the notion of continuous deformation. Quite simply, homotopy types are what you cannot change about a space any way you continuously deform it. And this way of describing what they are is good evidence that they possess some inherent simplicity that makes them suitable to play the role of the basic objects of a foundation of mathematics.

(C) “HoTT is merely a new kind of semantics for a pre-existing formal system. It was not purpose-built as a formal system in order to capture a pre-formal semantics in the same way that e.g. ZF(C) was purpose-built to capture a pre-formal understanding of the cumulative hierarchy of sets.”

Response to (C): It is true that HoTT builds upon a pre-existing formal system (intensional MLTT) that originally had nothing to do with homotopy theory. But the recycling and repurposing of formal systems is not only technically advantageous but also natural and unavoidable. Therefore, on pragmatic grounds, the fact that the original “meaning explanation” of Martin-Löf in [124] has nothing to do with the homotopic semantics of HoTT should not be seen as problematic. First-order predicate logic can also be seen as a repurposing of Frege’s calculus in the Begriffsschrift, but surely we wouldn’t want to claim that this presents conceptual worries for ZFC simply because of their differing intended semantics. Furthermore, in the specific case of intensional MLTT the higher structure of identity types was for a long time considered a kind of aesthetic flaw of the system since no-one could make sense of it semantically or find any use for it syntactically. Indeed, until Hofmann and Streicher [68] constructed the groupoid model

---

43 Marquis makes a very similar point in arguing for the fundamentality of homotopy types: “I submit that the notion of [homotopy equivalence] at work here is philosophically fundamental: we are dealing with entities that can be continuously transformed into one another.” ([121], p. 2151)
of MLTT it was considered plausible that the UIP axiom (which would collapse the higher structure of types and is incompatible with univalence) was provable in MLTT. Through the homotopy interpretation, this higher structure of identity types is once and for all revealed as a feature rather than a bug of intensional MLTT. And this, to my mind, suffices to quell any worries that intensional MLTT is being somehow misappropriated by UF.

2.5.2. Native formalizations for all theoretical contexts? In Section 2.4 I argued for (CI) by describing a general method for formalizing criteria of identity for mathematical structures as homotopy equivalences in UF. By virtue of this general description, this method is uniform in the sense of Section 2.2: wherever it applies, it applies for the same reason, i.e. for the reason that it encodes criteria of identity as componentwise isomorphisms. In addition to the objections considered above we must now ask whether the method is wide enough: does it plausibly cover all structures of mathematical interest? We must also ask whether it is native to mathematicians: does it yield formalized notions comprehensible to mathematicians familiar exclusively with their informal counterparts?

Let me begin with the first question, regarding the scope of the general method. There are two approaches one could take here, one formal, one informal. The formal approach would consist of providing a general notion of a signature, plausibly wide enough to cover all structures of mathematical interest, for which we would then prove a result analogous to the Structure Identity Principle in [171]. I believe it is possible to define such a general notion of a signature and have developed what I think is a plausible candidate in Chapter 5. Proving a Structure Identity Principle in terms of this general notion of signature is certainly conceivable. However, I do not think that establishing (SFOM) ought to rely on such a formal result. I understand (SFOM) here as an informal thesis, gaining in
generality what it perhaps loses in formal rigour. This is why I think it suffices for the purposes of this Chapter to rely solely on an informal argument for (F).

This informal argument would need to establish that most theoretical contexts in the practice of mathematics can be formalized using the general method outlined in Section 2.4.2. To do so, first note that informal criteria of identity will always strive to preserve all and only the structure of the objects under consideration. Insofar as we are clear on what that structure is, then there should be a clear and standard way to list the components that are relevant to the objects under consideration. For example, if we are interested in categories with specified limits, then we know that we should be able to separate their objects up to equality and not merely up to isomorphism. Second, as long as we are able to list these components, then we are able to formalize them inside UF, at least insofar as my responses to the objections in Section 2.5.1 hold up. This will give us a formal definition of $O_{UF}$ in HoTT. And then, finally, if we follow the method of Section 2.4 the correct criterion of identity will be phrased in terms of isomorphisms between these (formalized) components and will thus coincide with $\cong_O$. Therefore, insofar as a mathematical structure of interest can be broken down in components (each of which may come equipped with its own “local” criterion of identity) then the class of such structures can be invariantly formalized according to our general method.

However, this simple-sounding vision will face many difficult cases. Firstly, with respect to the “small” theoretical contexts like the natural numbers that are best formalized as individual types, we might find several examples in UF that strain the principle of regarding identities in them as degenerate versions of homotopy equivalence. An important such case is the synthetic formalization of homotopy types, e.g. of the circle $S^1$ as a higher inductive type. It is admittedly unclear how my method can make sense of such a formalization. In my defense, there is a blurry line between the role of higher inductive types as
formalizations of mathematical notions and their role as actual mathematical structures themselves. This line is likely to remain blurry until a general specification for higher inductive types is agreed upon. Until such time comes it is difficult to say much about how problematic such cases really will turn out to be for (SFOM).

Secondly, HoTT currently faces difficulties in formalizing notions from higher category theory (paradigmatically, those developed by Lurie in [99] and related work). In particular, somewhat ironically, HoTT is not very good at formalizing itself. Due to issues related to being able to express infinite so-called coherence conditions, the standard versions of HoTT (e.g. the one in [171]) are not able to express what it would be for a type (e.g. a universe $U$) to be a model of HoTT. Enriched problems arise for other such $\infty$-structures’. There is therefore a likelihood that as more of these structures find their way into the practice of mathematics, the less well HoTT will be able to live up to (SFOM). I think this is a serious issue to consider, but it is not one that an appropriately designed HoTT could not fix. Indeed, there exist solutions that have already been proposed (e.g. Voevodsky’s HTS) but they remain underdeveloped. But in any case, it must be conceded that until some general consensus emerges as to what kind of HoTTs will prove equal to the task, such theoretical contexts involving $\infty$-structures will present difficult cases.

Thirdly, with respect to the “large” theoretical contexts like category theory, one may worry that my general method works only because of some especially nice algebraic properties of (higher) categories and that it will break down in “less algebraic” theoretical contexts. Take for example as our theoretical context topological spaces up to homotopy understood in the traditional point-set-topological way – for ease of reference I will call

---

44For a precise description to the problem, as well as an illuminating introduction to possible methods that could solve it, cf. [163].
the objects *traditional* topological spaces and their criterion of identity *traditional* homotopy equivalence. In the manner outlined by my general method, we may now try to understand a traditional homotopy equivalence as a component-wise notion if we agree that the correct criterion of identity for points in a space $X$ is for them to be connected by a continuous path. Namely, for a given topological space $X$ and two points $x, x'$ in $X$ we define $x \sim_{\text{inf}} x'$ as the class of continuous paths between $x$ and $x'$. But now we are in a situation where the topology $\mathcal{T}$ of the space $X$ will not in general respect $\sim_{\text{inf}}$ since it is quite possible that there is an open set $U \in \mathcal{T}$ such that $x \in U$, $x' \notin U$ but $x \sim_{\text{inf}} x'$. Clearly, in such a situation we will have to violate either step (1) or step (3) of the general method. For we will either have to sacrifice talking about open sets, thus ignoring one component of the notion in question, or fail to assert that the (formalized) notion of continuous path can be equivalent to identity. (If it were, then any two points connected by a path would have to be contained in the same open sets, something which does not hold informally.)

What I want to say in response to such cases is summed up as follows: if the intended criterion of identity for one component $S_1$ of a structure is *not* respected by some other component $S_2$ (that may depend on $S_1$) then it is *not* a criterion of identity. To elaborate, it is helpful to consider a toy case. So consider a theoretical context where structures in $O_{\text{inf}}$ are given by (at least) two components $X$ and $S$ and we impose some criterion of identity $\sim_X$ on $X$. A formalization of this set-up in UF is given by the following data (where we use the same notation for $X$ and $S$ as their informal counterparts):

\[
X : U \\
S : X \to U \\
\sim : X \to X \to U
\]
In the example of traditional topological spaces, one can think of \( S \) as a propositional family expressing whether a given term is in a certain open set and \( x \sim y \) as the type of continuous paths between any two terms \( x, y : X \) (where “continuous path” is defined in terms of some notion of topology in terms of which \( S \) is also defined).

Now, on the one hand, it may be the case that we have two terms in one component related by the criterion of identity in that component that are explicitly separated by another component. In terms of the toy example, this means that we have \( a, b : X \) such that there is \( f : a \sim b \) but also \( S(a) \) and \( \neg S(b) \).\(^{45}\) I think it is clear that if one is considering a structure that contains certain objects that are separated by one component \( S \) of that structure, then no relation \( \sim \) defined in terms of any other component that relates objects that \( S \) separates can properly be called a criterion of identity. Any criterion of identity deserving of the name must be one that respects any other component of the structure under consideration. This indicates, for example, that we should not formalize traditional topological spaces up to traditional homotopy equivalence in HoTT as \( h \)-sets together with a family of “open” subtypes satisfying the required conditions. For the fact that we can define continuous paths in terms of open sets in mathematical practice should not make us forever beholden to these open sets just as the fact that pre-formally we define isomorphism in a category by separating objects up to equality shouldn’t mean we are forever beholden to equality of objects.

On the other hand, there may not necessarily be any component “disrespecting” a proposed criterion of identity for another component, but that criterion of identity may be too “loose” to induce equivalences between the rest of the components of the structure. For, ideally, we should want that if \( a \sim b \) then for any other component \( S \) that depends on \( a \) or \( b \) we have \( S(a) \cong S(b) \) (where \( \cong \) is the preferred criterion of

---

\(^{45}\)Here \( \neg S(b) \) can be thought of as notation for \( S(b) \to 0 \).
identity for entities of the kind that $S$ takes values in). For example, this is true in the case of categories because an isomorphism between two objects also induces a bijection between corresponding hom-sets.\textsuperscript{46} Yet we are perfectly well-able to imagine criteria of identity that do not induce equivalences on the rest of the components of the structure but for which it is not contradictory to assert a saturation axiom in the style of univalent categories. But this is not something that should trouble us: for by asserting a saturation axiom we do ensure that all other components of our structure end up respecting our chosen criterion of identity in the way desired – and without any loss of information.\textsuperscript{47}

The main difficulty shown up by the case of traditional topological spaces is that it is often unclear in mathematical practice how to “remove” certain pieces of structure in terms of which criteria of identity are expressed. For example, in the case of categories-up-to-equivalence we might as well “remove” the notion of equality of objects once we have decided that the criterion of identity for objects is given by isomorphism. This is easily done in this case since none of the already-given components (arrows, identities, composition) separate objects more finely than isomorphism. By contrast, there is no easy way to “remove” equality of points in a traditional topological space in favour of continuous paths between them since open sets may very well separate two such points. Nor is there any straightforward way to simply “remove” open sets since the very notion of a continuous path is defined in terms of them.

On the other hand, it is far from clear that the intuitive notion of a continuous path between points must be expressed this way, i.e. in terms of open sets that already

\textsuperscript{46}More precisely, for any other object $c$ an isomorphism $f : a \cong b$ induces a bijection $f^* : \text{Hom}(c, a) \cong \text{Hom}(c, b)$ by transporting morphisms along $f$ and similarly for $\text{Hom}(\cdot, c)$.

\textsuperscript{47}What the above-described situation certainly does prevent us from doing is a Rezk-completion-style construction. The reason that Rezk completion works in the case of categories is exactly because we can use a given isomorphism to induce a bijection (i.e. $h$-equivalence) between hom-sets. This means that isomorphisms “already” respect hom-sets. But we have no reason to expect this to be the case in general. So the Rezk completion construction certainly does not generalize to arbitrary structures, as of course one ought to expect.
separate distinct points. Indeed, the field of abstract homotopy theory is largely devoted to setting up the theoretical context of “topological spaces up to homotopy” in such a way that this does not occur, i.e. by “algebraicizing” them in a useful way. But it is unclear whether this process of “algebraicizing” can always take place – and even less clear that it can always result in a theoretical context that practicing mathematicians would find practically indistinguishable from the one we began with. Nevertheless, it seems to me that such difficult cases usually arise because of an inevitable bias towards thinking of the objects of mathematics in terms of sets, which thus tends to force us into making very many distinctions that might not prove necessary to faithfully formalizing the structures of interest. The example of traditional topological spaces is a case in point: although the point-set definition of a topological space makes perfect sense and can be studied as a mathematical object in its own right, it is by no account forced upon us when we want to study spaces up to traditional homotopy.48

Even so, the second question remains: are the formalized notions we end up with native ones? Will they be comprehensible to the practicing mathematician (insofar as any formal system is)? In one sense, the idea of formalizing anything as a structure on a homotopy type is likely to confuse practicing mathematicians far more than the idea of formalizing such structures in terms of sets. But this is a premature criticism at least insofar as it depends on the novelty of formalizing mathematical objects in terms of

48One is inevitably reminded of Grothendieck’s beautiful ruminations in his *Esquisse d’un Programme* [56]:

[When one tries to do topological geometry in the technical context of topological spaces, one is confronted at each step with spurious difficulties related to wild phenomena. [...] This situation, like so often already in the history of our science, simply reveals the almost insurmountable inertia of the mind, burdened by a heavy weight of conditioning, which makes it difficult to take a real look at a foundational question, thus at the context in which we live, breathe, work – accepting it, rather, as immutable data. [...] It is this again which explains why the rigid framework of general topology is patiently dragged along by generation after generation of topologists for whom “wildness” is a fatal necessity, rooted in the nature of things. (pp. 258-259)
homotopy types. It is certainly an idea that no-one had taken seriously pre-UF and which at this point still may appear mysterious. In his recent talk [180] Voevodsky emphasizes exactly this point:

The third component [of any foundational system] is a structure that enables humans to encode mathematical ideas in terms of the objects directly associated with [the structures that provide the semantics for the sentences of this language in terms of mental objects intuitively comprehensible to humans. In UF this third component], a way to encode general mathematical notions in terms of homotopy types, is based on the reversal of Grothendieck’s ideas considered in [[76]].

Both mathematically and philosophically, this is the deepest and least understood part of the story.

But from its novelty, it does not follow that it is an idea definitively incompatible with mathematical practice (or, at the very least, with the mathematical practice of today).

On the other hand, to go back to a point made in Section 2.2.1, the idea that any structure and property can be transferred along isomorphic objects is indeed very close to contemporary mathematical thinking, as has been emphasized numerous times in the extensive literature on categorical foundations of mathematics. In that respect, the formalized notions we end up with in UF are much closer to the way of thinking of practicing mathematicians. The details of the formalization and the formal system itself are certainly at the present moment in time not as widespread as set-theoretic formalizations in terms of ZF(C). But the fundamental idea of caring about mathematical structures only up to isomorphism – vastly generalized in HoTT – is certainly a defining trait of the mathematical practice of today. And insofar as the general method outlined here
depends crucially on formalizing this exact trait, then the essence of the formalizations themselves should prove native to practicing mathematicians, at least modulo the type-theoretic language barrier.

A related worry here is whether the concept of identity has been mutilated beyond recognition – whether, that is, it is at all plausible to expect practicing mathematicians to make sense of the idea that the notion of identity that they employ in practice is actually a version of homotopy equivalence and ought to be formalized as such. And if that is so, then mustn’t really the correlation between structuralism as an imperative to not distinguish isomorphic objects be separated from the imperative to remain faithful to mathematical practice? Now it is certainly the case, as mentioned above, that most mathematicians are likely to react with befuddlement if asked to evaluate the idea that the notion of identity they use in their daily practice (e.g. the equality between two natural numbers) is just a degenerate version of homotopy equivalence. And even those that are experts in sub-disciplines that favour this kind of “invariant” thinking (e.g. algebraists, category theorists) are likely to resist accepting this idea even if they come to see its practical benefits. I would say that the practice of mathematics is split between mathematicians that favour identity over invariance (i.e. would rather keep the intuitive notion of identity as a property even if it means giving up full invariance) and mathematicians that favour invariance over identity (i.e. who would be willing to reconsider their notion of identity if it gives them nice invariance properties). The split is certainly not even, but it is, I believe, a significant enough split to prevent any rightful claim that one of these two attitudes (invariance over identity vs. identity over invariance) is definitively the one that is faithful to practice.

\footnote{I owe this objection to John Burgess.}
So if it can be decisively established that practicing mathematicians will choose to retain an intuitive notion of identity understood only as a proposition (a reflexive, symmetric, transitive relation that either holds or does not hold, such as the notion of identity on which set theory relies) over the desire to stay faithful to the mantra that isomorphic objects should be indistinguishable then we must certainly give up hope that UF can both better capture the ideal of structuralist foundations and at the same time improve how faithful foundations are to mathematical practice. But I think the situation is not as clear-cut as that. I think the idea of having a language that does not distinguish between isomorphic objects has become very natural to many practicing mathematicians (certainly not to all and likely not even to the majority), to the point where such mathematicians have probably become willing to reconsider the primitive notion of identity they are working with insofar as it allows them to enjoy the kind of invariance properties we have been discussing. But, inevitably, facts about current and future trends in the practice of a discipline must remain in the realm of speculation – and I can offer no decisive reasons that the balance will tilt towards those who favour invariance over identity and those that do not. Nevertheless, there remains enough of a suggestion that such could be the case, that the claim that structuralist foundation like UF can do justice to the “reckless” structuralist ideal of making isomorphic objects indistinguishable while at the same time remaining faithful to practice cannot, as far as I am concerned, be entirely abandoned.

In conclusion, the general method I have offered cannot be taken as definitive. But neither, of course, is it meant to be. The process of formalization, after all, is not itself definitively formalizable. Insofar as formalizing mathematical structures and arguments still requires human intervention there will always be difficult cases, requiring patches

50My hunch is that it will.
and choices not reducible to general methods and principles. Ultimately, my argument for (F) – and therefore also for (SFOM) – can only rely on an abundance of clear-cut cases in which the general method does apply rather than on a purported clear-cut argument that all cases are covered.

Thus, what I have argued for in this Chapter should not (and cannot) be taken to be a once-and-for-all-argument that UF will always provide a way of formalizing mathematical structures of interest in an invariant manner. I would probably go so far as to say that the general method I have described in Section 2.4 does provide such a general method for formalizing most mathematical structures of current interest. But mathematical practice is changing and we may soon begin to study structures with properties that are just as difficult or unnatural to formalize in UF as higher-categorical structures in ZFC. Therefore, insofar as (SFOM) is tied to the mathematical practice of today (and of the immediate future) no claim that I make in this Chapter can exceed it. Nevertheless, even with this limitation in mind, there is certainly an important conclusion that emerges from the current argument: with respect to (SFOM), UF presents a quite significant improvement over (extensional or structural) set-theoretic foundations of mathematics.

Does this now mean: off with set theory’s head? Not at all. In fact, I think UF and set theory, especially ZFC, are not best understood as competing foundations. The relationship I envisage between ZFC and UF is essentially that of successive (but co-existing) conceptual schemes in the Quinean sense. Neurath’s famous (and somewhat overused) metaphor of the mariner is apt: UF is slowly rebuilding the ship of the foundations of mathematics using the material of set theory on which it still relies to stay afloat. For example, ZFC is used to build the simplicial model of UF and the construction of homotopy types in ZFC still guides our intuitions about what ought to be true of them. A pluralistic picture, I believe, is the most appropriate attitude to take here – one in
which set-theoretic foundations and UF establish a symbiotic relationship. I will spell out this picture, as well as explain its advantages and challenges, in Chapter 3.
CHAPTER 3

A Pluralistic Account of Foundations

In the preceding chapters I have been investigating the Univalent Foundations without asking: why do we need such foundations when we already have traditional and well-established proposals such as set theory? To answer, I will now investigate the way in which structuralist and traditional foundations could interact. Although much of what I say is focused exclusively on UF, the pluralistic picture I paint could certainly also apply to alternative structuralist foundations, such as the ones I will consider in Chapter 4. In Section 3.1 I will argue that traditional and structuralist foundations are not competing foundations because they are engaged in distinct problematics. Then in Section 3.2 I will argue for a pluralistic picture in the foundations of mathematics that makes sense of them as complementary foundations – roughly, structuralist foundations will assist the practical formalization of mathematics into proof assistants and traditional foundations will act as benchmarks of consistency. Then, in Section 3.3, I will examine the question of whether HoTT (as a formalization for UF) can receive a philosophical justification (a “meaning explanation”) independent of the justifications that have traditionally been given to set theory or type theory. Finally, in Section 3.4 I will explain how such an independent justification provides a definitive response to an objection raised by Hellman and Shapiro against categorical foundations of mathematics.¹

¹Early versions of Sections 3.1 and 3.2 were presented in a Symposium on the Foundations of Mathematics (“SoTFoM”) in Birkbeck College, London, January 2015.
3.1. A Non-Competitive Account

By a traditional foundation I will mean a foundational system developed in order to provide a philosophically justified formal system expressive enough to encode all mathematical arguments used in practice with the aim of providing a universal standard of rigour and a common language for all of mathematics in which constructions can be understood across disciplines and disputes objectively adjudicated.\(^2\) Given the developments in the 20\(^{th}\) century, set-theoretic foundations in the form of ZFC emerge as the dominant such traditional foundation. That said, constructive or intuitionistic variants of set theory (e.g. Aczel’s CZF [1]) and even alternative constructive foundations in general (e.g. MLTT prior to UF) I still consider to be traditional foundations. As such the divide between traditional and structuralist foundations mirrors neither the classical/non-classical nor the categorical/set-theoretic divide. Something being a categorical foundation does not automatically make it structuralist, nor does something being a foundation with a classical underlying logic make it traditional.

What defines a traditional foundation is that it is engaged in the problematic of trying to balance simplicity with expressive power. Traditional foundations aim to provide formal systems that are as expressive as possible while at the same time remaining as simple as possible so as to minimize the risk of inconsistency (a risk which Gödel’s second incompleteness theorem shows to be ineliminable at least in systems strong enough to serve as foundations for all of mathematics).\(^3\) More concretely put: the aim of traditional foundations is to find consistent, finitely axiomatizable theories into which the largest possible variety of mathematical theories can be interpreted.\(^4\) Since absolute consistency

\(^2\)For a recent account of the development of set theory faithful to this narrative, see [27].

\(^3\)I think it is uncontroversial to say that any formal system capable of serving as the foundation for all of mathematics will have to be strong enough to fall under the scope of Gödel’s theorems.

\(^4\)A nice way to put this point is in [141], p. 303: “[The goal of foundational studies] is to construct a consistent finitely axiomatizable theory whose type in the [interpretation lattice of theories] is as high as possible.”
is an impossible ideal, the demand for consistency in practice becomes a demand for the
greatest possible simplicity and philosophical clarity.

On the other hand, structuralist foundations aim to be optimal solutions to the prob-
lem of balancing faithfulness to mathematical practice with rigour. To be faithful to
mathematical practice means, in our case, to make no distinctions that practicing math-
ematicians themselves don’t make (e.g. between $2$ as $\{\emptyset\}$ or $\emptyset, \{\emptyset\}$). To be sure,
there is a trivial solution to this problem: one could merely use as such a foundation that
part of natural language that mathematicians regularly employ both in speaking and in
writing. But, of course, to do so is to abandon any claim to rigour. This is why there
has to be a balancing act. In trying to find a structuralist foundation for mathematics
one has to try to remain as faithful as possible to mathematical practice but without
violating some established standards of rigour. And a minimum such standard should
certainly be the requirement of being interpretable into (i.e. relatively consistent with)
some well-known and well-understood foundational system. A structuralist foundation
will therefore depend on some pre-existing (traditional or structuralist) Foundation that
can certify its consistency. In the current state of play, this means that any structuralist
foundation will depend on (some possible variant or reasonable strengthening) of ZFC.\footnote{It is exactly for these reasons that the first explicit construction of a model of UF (cf. [77]) was carried out in ZFC with two inaccessible cardinals and current work on a general theory of the syntax of UF (e.g. the series of papers of Voevodsky beginning with [179]) is being carried out in the setting of ZF set theory.}

This observation raises the crucial point. In addition to what said above, (formal
systems implementing) traditional foundations also play the role of what we may call
a \textit{benchmark of consistency}, i.e. a system that is judged fit to certify any other formal
system someone might be interested in studying. Thus, a “benchmark of consistency”
stands to mathematics in roughly the same relation that a universal Turing machine
stands to the theory of computation. In acting as benchmarks of consistency, traditional
foundations care little about how far they diverge from everyday mathematical language. (Similarly, we care very little about how practical it is to implement a universal Turing machine when we use it to prove theorems about computability.) On the other hand, structuralist foundations aim primarily at (SFOM), in the sense outlined in Chapter 2. Such an aim is clearly orthogonal to the aim of becoming a benchmark of consistency. Structuralist foundations, by their very definition, do not and cannot aim to usurp traditional foundations as benchmarks of consistency. Indeed – as is very clear in the case of UF – structuralist foundations require traditional foundations to certify them, before they can even enter the foundational fray as serious contenders.

To be sure, this is not to say that it is inconceivable for a structuralist foundation to attain the status of a benchmark of consistency. Perhaps many decades hence, a UF-based formal system will have endured the same kind of metamathematical gymnastics (standard/non-standard models, independence results etc.) that would allow it to obtain levels of foundational fitness comparable to those of ZFC. Perhaps then it too will be a worthy candidate to serve as a benchmark of consistency. But this is an entirely empirical matter that cannot be argued for in advance.

Importantly, this does not entail that UF is not a Foundation of mathematics. It entails that it is not a traditional Foundation of mathematics; it is, rather, a structuralist Foundation of mathematics. Structuralist and traditional foundations of mathematics are engaged in distinct problematics: balancing simplicity with expressive power on the one hand and rigour with faithfulness to practice on the other. Therefore, since traditional and structuralist Foundations have different goals, they are not competing proposals.

3.2. A Pluralistic Picture

From the fact that structuralist and traditional foundations do not compete with one another, it does not follow that they complement each other, or that there is any kind of
interesting pluralistic picture to place them both under. After all, even if UF is not in
competition with ZFC, why not view it as just another formal system definable inside –
and therefore part of – ZFC? What need do we have to add a structuralist Foundation
to our already well-established traditional Foundations?

To answer, we must look to mathematical practice. What we find there is something
not entirely uncommon to academia: overspecialization, insularity and a level of technical
sophistication that leads to great difficulties in efficiently and correctly refereeing new
results. In short, the complexity of modern mathematical techniques has outrun the
mathematical community’s capacity to verify their correctness.\textsuperscript{6} And the fact that ZFC
has provided a way to carry out such verification in principle does not help resolve
the practical problem of carrying it out efficiently. Translating submitted mathematical
papers into the formal language of ZFC would be akin to trying to certify the structural
integrity of a building by demolishing it, checking the molecular structure of every single
brick, and then rebuilding it from scratch. (The inhabitants will have settled in a new
house long before their old house becomes inhabitable again.) A structuralist foundation
is needed in order to improve this situation, i.e. in order to enable the formalization
of actual mathematical practice as it is presently taking place, with a view to making
it possible for new mathematical constructions and theorems to be proof-checked by
computers as they are being developed. You check the integrity of the house as you
are building it – so that, once built, you can move right in. In short, a structuralist
foundation is needed in order to turn the formalization of mathematical arguments from
philosophical idealization into practical reality.

\textsuperscript{6}This observation was Voevodsky’s original motivation in developing UF, prompted by the discovery of
certain errors in his paper \cite{76} that had remained undetected for several decades. For an autobiographical
account of this see \cite{180}. However, HoTT/UF has since acquired a life of its own and many of the people
currently involved in the project are not involved in it for these specific reasons. So it certainly has to
be made clear that HoTT/UF is not \textit{only} about developing usable formal proof-assistants, even though
that was one of the (many) reasons that motivated its development.
As I have pointed out in Chapter 1, the situation seems entirely analogous to the pre-Fregean state of affairs in the 19th century. In those days, many mathematicians were unbothered by the lack of rigour in mathematical arguments (especially in analysis) because they thought it legitimate to rely on intuition to fill in gaps in their reasoning. Similarly, in the present moment in time, mathematicians are largely unbothered by the fact that the currently accepted foundational languages of mathematics are too far removed from the practice of mathematics. Just as reliance on intuition allowed 19th century mathematicians to ignore the pressing problem of establishing a universal standard of rigour, so is reliance on formalizability (“It can all be formalized in ZFC if one is patient enough!”) allowing 21st century mathematicians to ignore the pressing problem of verifying obtained results. And just as the attempt to rectify the former issue led to deep and lasting insights into the nature of mathematics and philosophy, so, I believe, will the attempt to rectify the latter.

All this, however, raises an obvious question: Why can’t a traditional or structuralist foundation play the role of a benchmark of consistency and allow practically feasible formal encoding at the same time? I don’t think this is possible for the following reason. In order to achieve their goals, structuralist foundations will have to be synthetic theories of complex objects (e.g. ∞-groupoids). This is because structuralist foundations will need to postulate substantive invariance principles (e.g. the axiom of univalence). And such invariance principles can only be substantive if there is a lot of structure in the basic objects that they relate. To illustrate: if I say that I will not distinguish between individual grains of sand (my “invariance principle”) I am not saying anything substantive since individual grains of sand (my “basic objects”) are already practically indistinguishable. But if on the other hand I say that I will not distinguish between individual mounds of sand I am saying something quite substantive indeed, viz. that for my purposes one
mound of sand will do just as well as any other and that there is nothing that I can say of one mound that will not also be true of any other (I cannot say that “Timmy slid down this mound” unless Timmy slid down every mound). If on this sandy paradise of both Cantorian sand and Univalent mounds we think of mathematical constructions as sandcastles, then traditional foundations like set theory attempt to build them by assembling individual grains of sand in the right shape, whereas structuralist foundations build them by carving the right shape into a pre-existing mound of sand.\footnote{This distinction between “top-down” and “bottom-up” has been made many times in the literature on category theory, e.g. by Awodey \cite{awodey2004category}.} This methodological divide, it seems to me, cannot be fully bridged. This is why we need to make sense of structuralist and traditional foundations as entering a \textit{symbiotic} relationship.

Another – related – way to picture the distinction between structuralist and traditional foundations of mathematics is in the following terms. Every foundational framework $\mathcal{F}$ – in the sense of Chapter 2 – contains (at least) the following three components: a formal syntax that is interpretable in an intended semantics (which we understand as the “basic objects” of $\mathcal{F}$) together with an (informal) process of encoding what goes on in mathematical practice in terms of this intended semantics. Diagrammatically:

$$
\text{Syntax of } \mathcal{F} \xrightarrow{\text{interpretation}} \text{Semantics of } \mathcal{F} \xleftarrow{\text{encoding}} \text{Mathematics}
$$

In traditional foundations the main aim is to bring the Syntax as close as possible to the Semantics with little regard as to how far the Semantics end up from Mathematics. In other words, in traditional foundations the aim is to make \textit{interpretation} as easy as possible even at the cost of making \textit{encoding} very cumbersome. Thus, diagrammatically, we can picture set-theoretic foundations as follows (where $V$ is the “real” universe of sets):

$$
\text{ZFC} \xrightarrow{} V \xleftarrow{} \text{Mathematics}
$$
On the other hand, in structuralist foundations the aim is to bring the Semantics as close as possible to Mathematics with little regard as to how far they end up from Syntax. This is so that the encoding of mathematical arguments becomes as easy as possible, even if this comes at the cost of making the act of interpreting the Syntax into the Semantics very cumbersome. Thus, diagramatically, we can picture a structuralist foundation like UF as follows:

\[ \text{HoTT} \longrightarrow \infty \text{Gpd} \longleftarrow \text{Mathematics} \]

The difficulty of interpretation in structuralist foundations makes the metamathematics of its syntax far more difficult to carry out than in traditional foundations. This is one reason why traditional foundations are more well-suited to act as benchmarks of consistency. On the other hand, the ease with which encoding can take place in structuralist foundations makes them far more suitable to be implemented in proof assistants (that have a chance of being used by working mathematicians).

As the above hopefully makes clear, the need for structuralist foundations has not come about because of any philosophical deficiency with traditional foundations. From a purely philosophical standpoint the need for structuralist foundations may thus appear unmotivated. I agree: one has to look at the practice of mathematics in order to adequately motivate the development of structuralist foundations.

These difficulties are outlined in [163] and will also be discussed further in Chapter 4. Perhaps this means that structuralist foundations may not have anything new to say on the problems that have traditionally occupied the philosophy of mathematics, e.g. issues of the ontology of mathematical structures and the epistemology of mathematical propositions. But I do not find this possibility worrisome. As far as I am concerned, the deepest philosophical question that emerges out of this whole discussion is not “What can structuralist foundations do for the philosophy of mathematics?” but rather “Why is it that the need for structuralist foundations has now arisen?” Explaining why this need arose – driven by deficiencies in the communal practice of mathematics and especially the fact that the complexity of mathematical constructions has outrun the communal capacity to verify them rigorously – is, for me, the deep philosophical problem that lies in the background of this entire discussion. In order to even begin to answer it, however, we need to be very clear on exactly how traditional and structuralist foundations can live together, which is the question we are currently engaged in.
Thus, in order to clarify how this pluralistic picture is supposed to work let us look into HoTT and ZFC. The complementary relation between these two systems may be summarized diagrammatically as follows (where the length of the arrows is proportional to the difficulty of the task they represent):

![Diagram of ZFC and HoTT complementary relation]

The formal translations would be formal syntactic constructions of one system into the other. In the above picture then this would be something like the construction of the simplicial model of univalence [77] in ZFC with two inaccessible cardinals (hereinafter “ZFC+2U”) or the cubical model (cf. [21], Section 8) in one direction and the construction of a model of ZFC as a higher inductive type in HoTT (cf. [171], 10.5) in the other direction. The informal translations would be the ideas encoded by the formal translations, informally expressed. For instance, one possible informal translation of $\infty\text{Gpd}$ into $V$ would say: “An $\infty$-groupoid is a set of objects together with a set of arrows etc.” On the other hand, an informal translation of $V$ into $\infty\text{Gpd}$ would say that sets are merely the 0-types, i.e. those homotopy types such that identity types between any of their points are either empty or contractible.

With ZFC and HoTT as the blueprint, this then is how I propose structuralist and traditional foundations can co-exist. Structuralist foundations facilitate the encoding of currently-practiced mathematics into their formal language, thus (eventually) leading to proof-assistants that mathematicians may use to verify their arguments. Traditional foundations, on the other hand, facilitate the process of carrying out the metamathematics.
of such structuralist foundations. As the length of the arrows above suggests, it is much easier to translate the syntax of traditional foundations into its intended semantics (think of how natural it is to define set-theoretic semantics for first-order logic) and this is what makes traditional foundations a much better setting for metamathematical investigations. On the other hand, it is much more cumbersome to translate the syntax of structuralist foundations into its intended semantics (think of how mysterious it still sounds to say that types in intensional MLTT can be understood as homotopy types) but much easier to encode mathematical structures of interest in terms of their intended semantics. This is what makes structuralist foundations a much better formal system to have practicing mathematicians translate their arguments into.

Thus, in this pluralistic picture, Foundations are no longer to lie “behind the scenes” of mathematical practice, invisible to practicing mathematicians. Structuralist foundations like UF are designed explicitly not to stay behind the scenes (neither, of course, are they designed to take center stage). They are meant, rather, to remain in the foreground, visible to audience and performers alike. To push the theatrical simile (perhaps a little too far): mathematicians are the performers of the play; structuralist foundations are the stage designers that make sure the performance is comprehensible to the audience; and traditional foundations play the role of the engineers that guarantee the structural integrity of the building and the proper functioning of the stage (platforms, curtains etc.)¹⁰

Just like stage designers are constrained by the subject matter of the particular theatrical work being staged, so structuralist foundations are constrained by the actual practice (performance) of mathematics. Similarly, just as the engineers and superintendents of a stage usually have no need to know the specifics of the performance their stage is going to host unless some daring new demand is made of the stage itself (e.g. filling it with

¹⁰As for the audience, one can take that to be the referees, or even the philosophers, who are trying to convince themselves that what is being presented is a correct mathematical demonstration.
water or having horses gallop through it), so traditional foundations have no need to take into account specific mathematical techniques employed by practicing mathematicians unless some daring new demand is made of them (e.g. defining the category of all sets) in which case they briefly come to the foreground, tighten a few screws and add a few extra mechanical features to the stage (e.g. an extra Grothendieck universe of two) before once again retreating backstage and out of sight.

Importantly, stage-designers are in direct contact with engineers and performers alike – they understand both the artistic demands of the latter and the practical concerns of the former. In the mathematical theater, until recently, communication between engineers and performers was non-existent.\textsuperscript{11} The performers simply assumed that the stage was sturdy and the engineers made no particular effort to grasp the aesthetic value of the play being staged on it. Structuralist foundations, as stage designers, will help re-establish communication between the two sides, with benefits to both sides. Until recently, the foundations of mathematics cared only for the integrity of the building in which the play was staged. Supplementing the performance of mathematics with good stage design is what structuralist foundations, essentially, are about.

This is the pluralistic picture I am sketching: one in which mathematics no longer takes place on an empty (but sturdy) stage; one in which the stage design against which the mathematical theater is played out is once again deemed important to the performance itself. And just as the production of a particular play needs to consult closely with the stage on which it is to be produced, so must structuralist foundations work closely with the traditional foundations which certify them. Similarly, just as the engineers must be willing (occasionally) to acquiesce to the demands of ambitious stage design for the

\textsuperscript{11}Harvey Friedman, for example, has spent much of his career attempting to convince practicing mathematicians that there are are concrete mathematical statements independent of ZFC.
benefit of a production, so must traditional foundations be willing to acquiesce to the (e.g. proof-theoretic) demands of structuralist foundations.

This is how, in my view, structuralist and traditional foundations can work together in a complementary and not a competing fashion. This complementary relation amounts to a modified picture of what we should call a foundation of any discipline. To distinguish it from the way in which we have been using the term “foundation” so far, let me call this new notion a grand foundation. A grand foundation of a given discipline should consist of a foundation of that discipline’s practice as well as provide the ground on which the fundamental tenets of this discipline can receive some kind of “ultimate” justification (e.g. the way Russell wanted to found mathematics on the self-evidence of logic). In the case of mathematics, the development of set-theoretic foundations has had incredible (near-universal) success in providing the ground upon which mathematical inferences can be justified. Set theory has built an admirably sturdy playhouse around an extremely flexible stage (capable of hosting the widest possible variety of performances). But it has done so at the cost of alienating itself from mathematical practice. It has made the practice of mathematics itself appear (if not to mathematicians themselves then certainly to philosophers) as the theatrical equivalent of those austere postmodern productions that do away with all stage props and costumes and have a single performer, speaking in one voice, standing still at the center of the stage, enacting the entire play with a near-total disregard for context, plausibility or motion. Structuralist foundations enter this sturdy stage as production designers – they restore to the activity of mathematics its context, without taking away from the possibility of “ultimate” justification.

Of course, the pragmatic motivation for this pluralistic picture raises an obvious objection. If what we really care about is making encoding as easy as possible for practicing
mathematicians then surely we are not helping our cause by trying to make these mathematicians learn a language that is foreign to their (informal) way of talking and thinking. After all, whether we like it or not, mathematicians are used to set theory; they are used to, for example, thinking of functions as subsets of cartesian products of sets, not as terms of $\Pi$-types. To try to get them to change their way of thinking from set theory to type theory would surely defeat the point of trying to remain “faithful to practice” since so much of mathematical practice is carried out by casually employing “set-talk” (even if this does not involve conscious awareness of the axioms of ZFC).

There are three things to say in response. Firstly, the translation of set-theoretic arguments to type-theoretic ones is more straightforward than one would expect.\textsuperscript{12} Secondly, there \textit{is} a set theory inside UF (the theory of types of $h$-level 2, i.e. the $h$-sets) and so most mathematical arguments involving sets and structures on them can be translated verbatim, ignoring the “higher dimensions”.\textsuperscript{13} As such, mathematicians working in the most concrete areas of mathematics will not have to alter their language at all if they are to encode their arguments in UF. Thirdly, UF undoubtedly facilitates the formalization of several branches of abstract algebra widely used in contemporary mathematics and mathematical physics.\textsuperscript{14} Therefore, UF makes formalizing set-level mathematics no more difficult and makes formalizing higher-level mathematics easier. Overall, even modulo the language barrier, UF should make the formal encoding of mathematical arguments significantly more practical.

\textsuperscript{12}See [8] for evidence of that.

\textsuperscript{13}The caveat here is that the (1-)type of all sets in a univalent universe is naturally a model of ETCS, not of ZFC. See [171], Chapter 8.

\textsuperscript{14}Roughly, any algebra employing (higher) category theory, e.g. as in the work of Lurie [99]. But there is a clear caveat here: I am talking about \textit{direct} formalization of such higher algebra into HoTT. For those mathematicians used to working with higher algebra \textit{inside} set theory (e.g. $(\infty, 1)$-toposes as quasicategories) their arguments will have to be “elementarized” before they can be translated into HoTT. And that might not be a straightforward task.
But even if one is not convinced that this is true of the current practice of mathematics, there is nothing to say that it won’t be true of the future practice of mathematics. For monumental shifts in perspective do certainly occur in mathematics, and our present understanding of mathematical practice should not limit our imagination of what that practice may evolve into. As evidence of this, one need only recall the way in which Cantor’s set theory was perceived by mathematicians of his time, e.g. Kronecker is said to have remarked: “I don’t know what predominates in [Cantorian set theory] – philosophy or theology, but I am sure that there is no mathematics there.”\(^{15}\) And further, if the history of science has shown us anything, it is that radical overhauls of conceptual frameworks are indeed possible. To quote Manin once again, on the issue of moving from “Cantor sets” to homotopy types:

I am pretty strongly convinced that there is an ongoing reversal in the collective consciousness of mathematicians: the right hemispherical and homotopical picture of the world becomes the basic intuition, and if you want to get a discrete set, then you pass to the set of connected components of a space defined only up to homotopy. ([50], p. 1274)

Finally, it is important here to say a few words about how the pluralistic viewpoint sketched in this section compares to other “pluralisms” in the foundations and philosophy of mathematics. Firstly, note that the pluralism advocated here is between foundational proposals in the sense of Chapter 2, i.e. a pluralism of foundations according to the minimalist picture. This means that it is not a pluralism of languages nor a pluralism of formal theories. As such, what I have in mind has nothing to do with Feffermanian pluralism e.g. between predicativist and non-predicativist set theories or between standard and non-standard analysis (cf. [41]) or between classical and non-classical logic e.g.

\(^{15}\)This is a widely circulated, though seemingly apocryphal, quotation.
via “pluralities of toposes” (cf. [18, 65, 136]). It even has nothing to do with pluralisms that have been advocated between category-theoretic foundations (in the “old” sense) and set-theoretic foundations (cf. [63, 119]), or the set-theoretic pluralism [81] that is currently hotly debated among set theorists.

The pluralism advocated here, by contrast, is perhaps best described as a pluralism of discourses. I know of no close analogue to this view in the literature (this must be partly because the idea of two plausible foundational proposals with incompatible basic objects is very new) – but perhaps my view comes closest to the one expressed by McLarty in [134], where he argues that foundations are best understood as collections of truths that result in a useful and economical organization of mathematics. I would not go as far as to call the defining principles of each foundational proposal that I consider (e.g. univalence in the case of UF) truths – although McLarty’s view does not rely on any very substantive notion of truth. Furthermore, I would add to McLarty’s account the following requirement: Foundations need not merely be collections of organizing principles, but organizing principles for a particular purpose (or purposes). For example, for the purpose of providing a benchmark of consistency (as in the case of traditional foundations) or for the purpose of providing practically feasible formalizations (as in the case of (some) structuralist foundations). And in those terms the account presented here may then be expressed as follows: we should countenance many different ways of organizing mathematics, but also many reasons we have of doing so. It is in exactly this spirit that I think we can understand both UF and set theory as complementary proposals for the foundations of mathematics.

But in order to arrive at such a view, we need to broach the issue of autonomy. For if any kind of pluralism is to make sense, then we must ensure, it seems to me, that the two (or more) frameworks with respect to which we are being pluralists are actually
autonomous. And what does such autonomy boil down to in the case of foundations of mathematics? A useful characterization, originally meant for categorical vs. set-theoretic foundations, is given by Linnebo and Pettigrew [98], who say that it consists of the following three facts, which any foundation needs to possess if it can claim genuine independence from set theory:

- Logical autonomy: autonomy of formulation
- Conceptual autonomy: autonomy of understanding
- Justificatory autonomy: autonomy of motivation

With respect to the first two criteria, I think it is clear that UF and ZFC are autonomous: HoTT and ZFC are formulated over distinct basic languages (MLTT and FOL respectively) and can be understood independently one of the other, at least modulo sociological barriers such as exist between type theorists and logicians or homotopy theorists and set theorists. With respect to the third criterion, it also follows straightforwardly that UF is autonomous from ZFC insofar as we regard it as a structuralist foundation engaged in the problematic of balancing rigour with faithfulness to practice.\(^{16}\)

However, one important issue remains that is not included in the Linnebo-Pettigrew criteria. This is the issue of an intuitive (or pre-mathematical) justification that does not rely on (naive) set theory. What this amounts to, essentially, is the question of whether we can give a meaning explanation (to borrow Martin-Löf’s term) for HoTT. To this, the following section is devoted.

\(^{16}\)Interestingly, Linnebo and Pettigrew make a very similar point about the justificatory autonomy of categorical foundations by saying that such foundations ought to be motivated by a kind of naturalism, in which the objects are understood purely with respect to the utility they have for the practicing mathematician, i.e. purely in terms of what they *do* rather than what they *are*. This seems to be an intermediate form of structuralism, somewhere between the ontological (“mystical” or “hard-headed”) versions and the fully pragmatic version that I have adopted, following Burgess [27]. It would be interesting to explore its implications in the future and to what extent it is compatible with the picture painted by UF.
3.3. A Meaning Explanation for HoTT?

Phrases like “a pluralism of discourses” sure sound fancy, but they raise a further issue: how is each foundation justified at an intuitive level? Are traditional and structuralist foundations amenable to the same kind of justification? In particular, are the Univalent Foundations dependent on set theory and FOL= for their intuitive justification? Or is it possible to come up with an entirely autonomous intuitive justification for UF?

But before we get to all that, it is incumbent upon me to clarify exactly what I mean by an “intuitive justification”. To do so, it is helpful to explain how the demand for a justification of the rules of a deductive system differs in the case of a foundational formal system (like HoTT or ZFC) and a formal system meant for model theory (e.g. plain first-order logic or the n-logic of Chapter 5). On the one hand, in the case of a formal system meant for model theory, there is a perfectly precise sense in which one can prove a completeness theorem: one defines a class of models with respect to (or, rather, inside) a given foundational system and then checks that provability coincides with validity with respect to that class of models. Kripke’s completeness theorem for intuitionistic logic, for example, is such a result and so is (though less clearly) Gödel’s completeness theorem for classical first-order logic. Kripke’s completeness theorem proceeds not by justifying the rules of intuitionistic logic by giving them a pre-formal meaning, but, rather, it furnishes them with a justification through a formal semantics that takes place in set theory.\(^\text{17}\)

In the case of a foundational formal system (i.e. one that can serve as component (2) in the minimalist picture) there can be no completeness theorem in the formal sense since a foundational formal system is meant to describe the very “universe of objects” that will be used to define what a model is. Although of course we can talk about a model of ZFC (and can do so formally if we work in some appropriate ambient formal system) it

\(^\text{17}\)As has subsequently been shown (e.g. in [138]) these can also be defined in certain toposes by the so-called Kripke-Joyal semantics.
makes little sense to talk about a formal “completeness theorem” for ZFC. Rather, we can provide a “meaning explanation” for the axioms of ZFC and, exactly analogously, for the rules of (extensional) MLTT. This pre-mathematical “meaning explanation” provides a justification of these formal systems in terms of informal, primitive, intuitive notions. In other words, a meaning explanation, as I will use the term, is a philosophical (non-formal) completeness theorem. An almost endless variety of meaning explanations have been given for “core” ZF(C) set theory ever since its inception and also for plausible “large cardinal” extensions (cf. [105, 106]); the locus classicus in the case of dependent type theory remains Martin-Löf’s meaning explanations as outlined in [126] and [123] as well as the perspective taken by Dybjer in [38].

This difference between model theory and the meaning explanations for foundational formal systems is brought out very clearly by Martin-Löf in [125]:

[Model Theory] and translation into another language are one and the same thing [...] [Therefore] you can give meaning in this way only if you have another language to translate into. [...] But, eventually, you will of course have to have a language which is not given meaning by translating it into another language but has to be given meaning in some other way, and this is the language of the most primitive notions that you are dealing with, because that they are primitive means precisely that they cannot be defined in terms of any other notions. ([125], p. 408)

By “intuitive justification” I will therefore understand exactly this interpretation of the (rules of the) formal system (in this case, (a) HoTT) in terms of a language primitive enough not to require formalization into a pre-existing formal system (though certainly

18Indeed, this very point has been used by Mayberry in a series of papers to argue for a return to a naive (non-axiomatic) set theory in the spirit of Cantor, see [127–129].
such a formalization may be given, if one so wishes). Another way of explaining this
difference is in terms of “pre-mathematical” versus “meta-mathematical” justifications
for a formal system. A pre-mathematical justification is one that does not rely on
an interpretation of the given formal system within some equally formal metatheory,
whereas a meta-mathematical justification is one that interprets the given formal system
in some other formal system. To use Dybjer’s [37] example: a realizability model of
MLTT is a meta-mathematical justification, whereas the “testing semantics” of Dybjer
constitute a pre-mathematical justification. In this Section, I am also interested in a
pre-mathematical, rather than a meta-mathematical, justification of HoTT.

Thus, the question I want to tackle here is the following: What, if any, is the jus-
tification of the validity of the rules of HoTT? Or, more concisely: can there be a pre-
mathematical meaning explanation (in the sense of Martin-Löf) for HoTT? And if so, is
it fully independent of previously-given set-theoretic or type-theoretic meaning explana-
tions? I am in agreement with Martin-Löf here that such an endeavor, of providing a
meaning explanation instead of a (formal) semantics is

[a] genuinely semantical or meaning theoretical investigation, which
means that [one] must enter on something that [one] is not prepared
for as a mathematical logician, whether model theorist or proof theo-
rist: [one] must enter on an enterprise which is essentially philosophical
or phenomenological [...] in nature. ([125], p. 408)

and also when he writes that

[t]here are [...] certain limits to what verbal explanations can do when it
comes to justifying axioms and rules of inference. In the end, everybody
must understand for himself. ([123], p. 166)
Thus what I am after here is neither the kind of interpretation of HoTT in set theory achieved in [21, 77], nor the formalization of the “Initiality Conjecture” outlined in [179], nor even more recent efforts by Angiuli, Harper and Wilson [5] to provide computational meaning explanations in terms of realizability semantics – what I am after is a genuinely “meaning-theoretical” explanation of HoTT, one that, limited though it may be, opens up the possibility that anyone could sit down and understand the validity of its rules for him/herself, without having to rely on a lot of heavy machinery in algebraic topology or theoretical computer science already developed inside set theory or topos theory.

Let me first make a few remarks that seem to me fairly obvious and uncontroversial. Firstly, it is not satisfactory to rely on an intended model built in set theory (e.g. the simplicial set [77] or cubical set [21] models) and then “piggyback” on any meaning explanation one can provide for set theory. In other words, it is not a satisfactory answer to use as justification for the rules of HoTT the fact that these rules come out true when interpreted in set theory. Of course, this is not to say that such results are not interesting, even essential. What they do is provide formal justifications, i.e. relative consistency results; and by showing that HoTT is interpretable into ZFC+2U, for example, we certainly have more reason to believe in the consistency of the system, indeed at least as much reason as we have to believe in the consistency of ZFC+2U. But what we are after here is an intuitive, not a formal, justification.

Secondly, a meaning explanation for HoTT cannot be an extension of Martin-Löf’s original meaning explanation. His meaning explanation, for one, justifies the principle of Uniqueness of Identity Proofs (UIP) and UIP is inconsistent with univalence.\footnote{The justification for UIP goes roughly as follows: insofar as we have a proof of identity, we no longer care to distinguish between the two terms shown to be identical; therefore all that matters is that they are identical, which means that there should be only one canonical term demonstrating this fact. See [38], p. 8 for further discussion.} Thus the grounds upon which the additional rules of HoTT are to be justified cannot be an
extension of the grounds upon which the original rules were justified. Of course, this leaves open the possibility of ignoring the meaning explanation of rules like UIP that are inconsistent with the new rules and trying to extend what is left. But I do not think that this can work. After all, a justification for a rule is called a justification for a reason. Whether or not we choose to ignore this reason, it is still there, and it remains the same for all other rules. For example, the reason we give to justify the axioms of ZFC is that we have a certain conception of the cumulative hierarchy. If we choose to ignore one of the axioms and add another that is inconsistent with the one we removed then the reason we have for justifying this new system cannot be the same as the old system even if the justification may look superficially the same. What has happened, rather, is that our conception of the cumulative hierarchy has either changed, or we are no longer relying on a certain conception of the cumulative hierarch as the reason for our justifications at all. In short: incompatible rules can only be justified by incompatible reasons. Since there are incompatible rules in HoTT and extensional MLTT, any reasons we have for justifying either of them will be incompatible.  

Therefore, the only option left to us is that of providing a completely new meaning explanation for the rules of HoTT, one that neither piggybacks on set theory, nor extends Martin-Löf’s programming-inspired meaning explanation. Furthermore, let me note that this also means that MLTT and HoTT should be viewed as fundamentally distinct formal systems and not species of the same genus (a genus one may broadly label “dependent

---

And indeed, this seems to be something that Martin-Löf has also recognized to be the case:

> The informal, or intuitive, semantics of type theory makes it evident that closed expressions of ground type evaluate to head normal form, whereas metamathematics [...] is currently needed to show that expressions which are open or of higher type can be reduced to normal form. The question to be discussed is: Would it be possible to modify the informal semantics in such a way that it becomes evident that all expressions, also those that are open or of higher type, can be reduced to full normal form? (quoted in [38], p. 21)

In other words, Martin-Löf here seems to acknowledge that a new kind of “modified” informal semantics seem to be required for the justification of the intensional variant of type theory (on which HoTT is inevitably based) since those currently available are incapable of dealing with “higher types”, by which I understand the higher identity types of a given type.
type theories”). Although sociologically inevitable, to call HoTT a type theory is, I think, conceptually a mistake. Any formalization of UF should rather be called “∞-groupoid theory” in much the same way that ZFC (or CZF, or ZF etc.) is called a “set theory”. Dependent Type Theories, for me, are whatever formal systems can be given a Martin-Löf-style meaning explanation – and as I have argued above this is not true of (any) HoTT.

3.3.1. A meaning explanation for which HoTT?. Before we embark on the project of providing a meaning explanation for HoTT, we need to agree on a scope for this meaning explanation. Namely, for how strong a system of HoTT should such a meaning explanation be provided? A reasonable minimum scope seems to me the following: a meaning explanation should be provided for a fragment of HoTT that is able to interpret Peano Arithmetic (PA) or, if we wish to steer clear of excluded middle, Heyting Arithmetic (HA). For insofar as most “concrete” mathematics can be encoded in PA, any fragment of HoTT into which PA interprets would, if supplied with a meaning explanation, provide an intuitively justified formal system that can encode most mathematics that is of interest in practice. Thus, the project for a meaning explanation for HoTT should consist of the following three steps:

(1) Define a fragment of HoTT into which PA can be interpreted.

(2) Provide a meaning explanation for that fragment of HoTT that is independent from the meaning explanation of Martin-Löf.

(3) Argue that the meaning explanation in step (2) is independent of first-order logic and set theory.

21Exactly how to understand “strong” in the setting of dependent type theories is itself a complicated issue, cf. the proof-theoretic analyses in [161] and [55] for an illustration of the relevant techniques.

22Though much of this encoding will fail to be native in the sense of Chapter 2.
A couple of clarifications are in order. Firstly, in step (1), “can be interpreted” should be understood in the propositions-as-types manner, e.g. as sketched in [10] and also as summarized in Section 3.8 of [171]. Namely, for a given first-order sentence, we translate atomic formulas as certain type families and logical connectives and quantifiers as certain type constructors (e.g. \( \lor \) as \(+\), \( \exists \) as \( \Sigma \) etc.) To be precise, we use the following definition:

**Definition 3.3.1.** Let \( \phi \) be a first-order sentence, \([\phi]\) its proposition-as-types translation and \( T \) a type theory that extends basic MLTT to a sufficient extent that allows the \([\cdot]\)-translation to be well-defined. We say that \( \phi \) *interprets into* \( T \) if \( T \) can derive a closed term of type \([\phi]\), i.e. if a judgement

\[
\emptyset \vdash p : [\phi]
\]

is derivable in \( T \). A first-order theory \( T \) interprets into \( T \) if all its axioms interpret into \( T \).

In the long run, we should not necessarily restrict ourselves to a propositions-as-types (PAT) interpretation of PA, and there could be alternative (exotic) schemes that may prove useful. But, for now, PAT interpretations provide the best-understood method for carrying out interpretations of first-order logic into MLTT, and they will be the ones that I will employ here.

The first thing to note is that in the setting of MLTT, in order to interpret PA (or HA) in the sense of Definition 3.3.1, at least a universe is required. This follows from the much more general fact that no propositional disequality can be deduced in MLTT without a universe.\(^{23}\) So the HoTT in question will have to contain a universe. In practice this is an inessential restriction since a universe is anyway required in order to even be able to state

\(^{23}\)This fact was first proved in [166] and a nice explanation can also be found in [172], pp. 264-265.
the axiom of univalence (for the universe in question) and any formal system that does not contain a univalent universe hardly deserves to be called “HoTT.” 24 Secondly, the axioms of PA/HA make use of the full expressive power of FOL=, including negation for the fourth postulate (“0 ≠ Sn”). This means that we will require all the corresponding type constructors in MLTT (Π, Σ, +, Id, 0, 1). Finally, we will clearly require the natural numbers N defined as an inductive type. Thus, a “fragment of HoTT” minimally sufficient for our purposes contains the following:

(1) The usual syntax of MLTT consisting of the four forms of judgment.

(2) The usual structural rules of MLTT governing contexts, substitution and judgmental equality.

(3) Rules for Σ, Π, Id, +, 0, 1 and N.

(4) A universe U closed under all the above type constructors satisfying univalence.

For the purposes of this section I will abbreviate this system as 1UHoTT (for “one universe HoTT”). So from now on, “a meaning explanation for HoTT” is to be understood as a “meaning explanation for 1UHoTT.” 25 Clearly this is not the only system that adequately formalizes UF, nor is it clearly the most suitable one to receive a meaning explanation. Indeed, as I will discuss below, Cubical Type Theory (CTT) might prove better-suited to this task. Furthermore, 1UHoTT is a rather weak system, since in the absence of W-types it is unable to encode standard inductive definitions, as well as being essentially predicative in not including any “collapsing” or “resizing” axioms for the universe U.

24 It must be made clear, however, that even very “tiny” universes (e.g. Prop) suffice for the interpretation of PA (or HA) into type theory. Such “tiny” universes are much less structured than the universes of types of e.g. the HoTT system in [171]. But such “tiny” universes do not suffice for the purposes of stating univalence and so we cannot consider them as viable options here.

25 Formally, UniMath [181] is the closest formal realization of 1UHoTT, at least insofar as one ignores the fact that it is built on top of Coq, which is itself based on the much stronger Calculus of Inductive Constructions. Alternatively, 1UHoTT can be taken as the system described in the Appendix of [77] if we replace the rules of W-types with the rules for N (which of course can be seen as an instance of a W-type).
similar to the Propositional Resizing considered in [171] (Axiom 3.5.5). But for our purposes here, this is not a direct concern: we are looking for the weakest possible system that can interpret PA and be plausibly regarded as a HoTT. In any case, for now, 1UHoTT is the system we will work with here.

3.3.2. Towards a Spatial Meaning Explanation for HoTT. The main idea of my meaning explanation is to regard the rules of HoTT as instructions for constructing new shapes from old. A little more, but not fully, precisely: the meaning of type constructors and term constructors consists in describing what can be done on previously constructed shapes and their parts ("points"). For example, $\Sigma$-types are to be viewed as "gluing" shapes together and $\mathbb{N}$ is to be viewed as the infinite "discrete" shape. The rules of the system are then justified by arguing that moving from a given construction (or, more precisely, observation of a construction) to a new one is correct. And what is correct is determined by our spatial intuition, at which point we turn the spade.

So what is now required is, firstly, to use this idea to supply the basic syntax with meaning and then, secondly, to use this meaning to justify the rules of the type theory. I will begin with the first task. There are three basic elements of the syntax whose meaning we need to explain: judgments, hypothetical judgments and substitution. I will take them in turn.

Interestingly, UniMath does actually employ a very strong impredicative principle, namely $\mathcal{U}: \mathcal{U}$, which as has been shown by Girard [51] actually leads to an inconsistent system. In other words, for practical purposes UniMath works with an inconsistent formal system – which might sound like anathema to the classical ear, but is certainly a conceivable state of affairs in the setting of constructive mathematics, in which one cares primarily about extracting algorithms. An interesting possibility is raised whether this inconsistency – namely the assertion $\mathcal{U}: \mathcal{U}$ – is actually not intuitively incoherent in terms of the below given meaning explanation. For this statement asserts, as we shall see, that $\mathcal{U}$ can be observed to be a point on itself. Such a statement, to my mind, does leave open the possibility that a meaning explanation could be given to 1UHoTT supplemented by a principle such as $\mathcal{U}: \mathcal{U}$ that actually makes it inconsistent (though far more useful). But such an admittedly odd scenario will not be explored any further here.
As may be clear already, what I have above called “observations” refer to the “judg-
ments” of type theory.\textsuperscript{27} So let me begin by giving the general meaning of the four types of
judgments, which I will from now on call observations. In standard MLTT (and therefore
also in 1UHoTT) we have the following four observations, where I have written “\textit{Shape}”
instead of the usual “\textit{Type}” to hint at the intended semantics:\textsuperscript{28}

\begin{enumerate}
\item $\Gamma \vdash a : A$
\item $\Gamma \vdash a = a' : A$
\item $\Gamma \vdash A \text{ Shape}$
\item $\Gamma \vdash A = A' \text{ Shape}$
\end{enumerate}

Strictly speaking the “judgments” themselves are given by what is on the RHS of the
turnstile – but I will use the term “observation” to refer to the whole assortment of
symbols $\Gamma \vdash J$ (for $J$ any of the four judgements in the usual sense) in order to avoid
maintaining a “logical” reading of the turnstile. Thus, the meaning of the four kinds of
observation is then given as follows:

\begin{enumerate}
\item From \textbf{viewpoint} $\Gamma$, $a$ is a \textbf{point} of $A$
\item From \textbf{viewpoint} $\Gamma$, $a$ and $a'$ are \textbf{symmetric} as points of $A$
\item From \textbf{viewpoint} $\Gamma$, $A$ is a \textbf{shape}
\item From \textbf{viewpoint} $\Gamma$, $A$ and $A'$ are \textbf{symmetric} as shapes
\end{enumerate}

\textsuperscript{27}There is an unfortunate terminological clash with \textit{Observational} Type Theory (OTT), a variant of
MLTT considered in [4]. The terminological connection between my use of “observation” here and the
use of the same word in OTT is no more than a coincidence.

\textsuperscript{28}In 1UHoTT I would ideally rewrite these as follows:

\begin{enumerate}
\item $\Gamma \odot a : A$
\item $\Gamma \odot a : A \equiv a' : A$
\item $\Gamma \odot A \text{ Shape}$
\item $\Gamma \odot A \equiv A' \text{ Shape}$
\end{enumerate}

with “$\odot$” as a symbol meant to convey “sees” (or “observes” or “surveys”) and with (2) meant to convey
coincidence (or congruence) of points after some symmetry. Lest such new notation prove unnecessarily
alienating, I will retain the old notation, even if it is meant to be understood in the completely new way
sketched below.
There are thus four new basic notions, on which the whole proposed meaning explanation for 1UHoTT is supposed to rest:

(a) Viewpoint (replacing “context”)
(b) Point (replacing “term”)
(c) Shape (replacing “type”)
(d) Symmetric (replacing “judgmentally equal”)

I claim that these four basic notions (a)-(d) are intuitive enough to justify basing a meaning explanation upon them. Inevitably, it is some kind of geometric (or “spatial”) intuition that is here invoked. We can argue about the content of this intuition all we like – but, for now, all that is required is an agreement that the four notions (a)-(d) are sufficiently primitive and at the same time sufficiently evocative to serve their role in a pre-formal meaning explanation.

I take this to be uncontroversial, but I will briefly run through the four concepts (a)-(d) in order to eliminate as much ambiguity as possible. By “viewpoint” I understand a certain vantage point from which shapes, points, symmetries can be observed. Exactly how this works spatially will become clearer when I explain hypothetical judgments below. By a “point”, I understand the usual notion of an irreducible part in the standard Euclidean sense (“that which has no parts”). By a “shape” I understand arrangements of points possibly connected by paths. Finally, by “symmetry” I understand the following: in the case of shapes, two shapes are symmetric if there is a way to place one on top of the other in order to render them indistinguishable (for example, two pentagons that are out of alignment); in the case of points, two points are symmetric (or “congruent”) if

Formally, of course, “shapes” will be interpreted as ∞-groupoids, so a reader familiar with the way in which the latter define a notion of an “abstract shape” is more than welcome to take my intuitive shapes as encoding just this intuition. For more about this topic, see [165]. But this intuition is not required to make sense of the meaning explanation below. Also note that the notion of a “path” has been smuggled in here without having been previously introduced. The notion will be introduced below, when we talk about identity types.

155
they coincide after applying some symmetry of the shape in which they lie (for example, two points lying on opposite edges of a square, which coincide if we flip the square along the appropriate diagonal of symmetry).\footnote{It might be advisable to have two separate notions of symmetry, one for points (congruence) and one for shapes (symmetry). I have opted against this on grounds of parsimony – but the concerned reader should note that nothing I say below relies essentially on maintaining a single notion of symmetry, to be applied to points and shapes alike.}

Thus, to summarize, if the meaning of a judgment \( J \) in MLTT (under the traditional meaning explanation) is “to know that \( J \)”, the meaning of an observation in HoTT (under the spatial meaning explanation) is “to observe that \( J \)”, where now the \( J \) are interpreted in such a way that it makes intuitive sense to regard them as things one can observe (spatially). For example, if \( J \) is \( a: A \) then instead of taking this to mean that “we know that \( a \) is a program matching the specification given by \( A \)” we take it to mean that “we observe that \( a \) is a point of the shape \( A \)”.

Given this, we must now give an account of “hypothetical observations”, i.e. of judgements in traditional MLTT of the form \( \Gamma \vdash x: A \) with a non-trivial context on the left-hand side of the turnstile. These “hypothetical judgments” will be understood as “constricted observations” with the intuition being that an observation is made of a shape/point/symmetry from a specific (constricting) viewpoint represented by the points and shapes on the left-hand side of the turnstile. So, for example, take the observation

\[
x_1: A_1, x_2: A_2(x_1) \vdash a: A
\]

This is to be understood as saying: From any point \( x_2 \) observable from \( x_1 \) we can observe a point \( a \) of \( A \). And similarly for symmetries of points, shapes and symmetries of shapes.

To illustrate the visual intuition behind these “hypothetical observations” let us consider the following observation

\[
x: A \vdash b(x): B(x)
\]
As indicated above, $A$ and $B$ are shapes. The meaning of this observation is now the following:

“From any point $x$ of $A$, we can observe a point $b(x)$ of $B(x)$.”

Visually, one can think of it as follows: $A$ is a plane lying below a sphere $B$; from any point $x$ of $A$ one can see a certain portion of the sphere, namely the surface of the sphere that is observable from the “cone” that originates from $x$, which we denote by $B(x)$; and on this portion of the sphere, we are told, we can observe a certain point $b(x)$. Pictorially, the observation describes the following situation:

![Figure 1](image)

**Figure 1.** A shape $B$ visible from a shape $A$. The point $b$ is visible from the point $x$ of $A$.

And the idea is that if this picture in Figure 1 is twisted so that we now view the sphere from the point of view of $x$, the portion of the sphere that is visible contains a point, given by $b(x)$. This is illustrated in Figure 2.

Thus, the meaning of type dependency (or, if you prefer, of a *fibration*), in our setting, becomes the following: a type $B$ dependent on another type $A$ is to be understood as a shape patched together from the different portions $B(x)$ that we can observe through
Figure 2. The (circle-shaped) portion $B(x)$ visible from the point $x$ of $A$. The point $b(x)$ now lies on $B(x)$.

arbitrary points $x$ on $A$. As hopefully suggested by the above pictures, the mental image here should be one where from an arbitrary point $x$ on the plane $A$ we shine a spotlight on the sphere $B$ – and $B(x)$ is the circle that we can now observe on the surface of $B$.

The third feature of the raw syntax of MLTT that requires spatial interpretation is the notion of substitution. Here, things are more straightforward. Given a certain point $a$ on a shape $A$, and any expression $B$ containing a variable $x$ of type $A$, the expression $B[a/x]$ is to be understood as the expression (whether shape or point) in which we have fixed the specific point $a$ in place of arbitrary points $x$ on $A$. Since expressions will be parts of judgements (which I want to understand as observations) substituting specific points inside an expression is to be understood as “fixing” or “constricting” a viewpoint down to a single point. As we will see below, this basic idea suffices to justify the structural rules of 1UHoTT that involve substitution.

Before moving on to the rules, it is worth noting the absence of any mention of truth, or of similar notions usually associated with judgments and propositions (e.g. knowledge how/knowledge that as in [126]). This is both intentional and, to a certain
extent, inevitable. The “observations” of the four basic forms (1)-(4) above are best understood as fitting into a Euclidean framework: one sets up a *problem* and then is asked to carry out a certain *construction* (rather than, say, demonstrating the *existence* of such a construction, which would turn what is sought into a proposition with a truth value). Thus, to ask whether an observation is derivable is akin to setting up a geometric problem, and producing a derivation of this observation is akin to constructing a solution to this problem. For example, to ask whether \( \Gamma \vdash a : A \) is derivable is akin to asking whether a point can be constructed on the shape \( A \) as observed through \( \Gamma \) – and to produce a derivation of \( \Gamma \vdash a : A \) is then a construction of such a point. From this point of view, a notion of truth can then be recovered as a degenerate notion of construction, namely when we ignore all details of the construction itself and retain only the information that it is possible to carry it out.

### 3.3.3. The Justification of the Structural Rules.

Given this basic interpretation of judgements-as-observations, of hypothetical judgements-as-constricted observations and of substitution-as-fixing-a-viewpoint, I will now move to the justification the structural rules of 1UHoTT, i.e. the usual structural rules of MLTT.

This proceeds more or less unproblematically. For example, the rules for judgmental equality are justified by the fact that symmetry is clearly intuitively an equivalence relation, and that it “respects typing” because if a shape can be placed one on top of the other, then any point on either shape can also be seen as a point on the other shape. For example, take the following rule

\[
\Gamma \vdash a : A \quad \Gamma \vdash A = B \quad \text{Shape} \quad \text{Sym-tran} \\
\Gamma \vdash a : B
\]
What we have above the line is the assurance that we can observe a point \( a \) on a shape \( A \) and that shape \( A \) is symmetric to shape \( B \). Being symmetric means that these two shapes are spatially indistinguishable – think of \( A \) as a square and \( B \) as the square produced by flipping \( A \) along a line of symmetry. Then, clearly, the point \( a \) on \( A \) can also be observed to lie on \( B \), thus giving us the observation below the line, and thus justifying \textbf{Sym-tran}. The justification of the rest of the structural rules concerning judgmental equality proceed very similarly.

The justification of \textbf{Vble} is immediate: if we can observe a point, then we can observe the same point. The rules for weakening (\textbf{Wkg} \(_1\) and \textbf{Wkg} \(_2\)) and substitution (\textbf{Subst} \(_1\) and \textbf{Subst} \(_2\)) are also straightforward to justify, along similar lines. For \textbf{Subst} \(_1\) we need to justify the following rule:

\[
\Gamma \vdash a: A \quad \Gamma, x: A, \Delta \vdash b: B \\
\Gamma, \Delta[a/x] \vdash b[a/x]: B[a/x]
\]

What we have above the line is the assurance that we can observe a point \( a \) on a shape \( A \) and that from a viewpoint that includes an arbitrary point on \( A \) we can observe a point \( b \) on \( B \). Below the line we have the assurance that we can observe a point \( b[a/x] \) on \( B[a/x] \) from the same viewpoint as above, except instead of an arbitrary point on \( A \) we now use the specific point \( a \) that we already know we can observe. Clearly, if we fix a viewpoint on a specific point, then every observation we could make from an arbitrary point we can also make from that specific point too, thus justifying the rule in question. The justification for \textbf{Subst} \(_2\) proceeds similarly.
For Wkg$_1$ we need to justify the following rule:

\[
\Gamma \vdash A \text{ Shape} \quad \Gamma, \Delta \vdash b : B \\
\frac{}{\Gamma, x : A, \Delta \vdash b : B} \text{ Wkg$_1$}
\]

What we have above the line is the assurance that we can observe a shape $A$ from viewpoint $\Gamma$ and a point $b$ from viewpoint $\Gamma, \Delta$. Now since the point $b$ is observable from $\Gamma, \Delta$ it is certainly also observable from $\Gamma, x : A, \Delta$ since, on the assumption that this latter context is well-formed, the observation of $b$ will proceed just as before. In other words, we already know what we need to observe $b$, and as long as an arbitrary point of $A$ is observable from $\Gamma$, the rest of the points we need to observe $b$ are already there, and the addition of $x$ doesn’t change the situation at all. The justification of Wkg$_2$ proceeds similarly.

Finally, as usual, we also have structural rules for each specific type constructor ($\Pi$, $\Sigma$, etc.) asserting that the conclusions of the rules are invariant under judgmental equality. These can be straightforwardly justified assuming the justification of the rules of each specific such type constructor, given below.

### 3.3.4. The Justification of the Rules for Type Constructors.

With this in mind, we must now provide the spatial meaning explanation for the rules of specific type constructors. In the case of 1UHoTT, these type constructors are the following: $\Sigma$, $\Pi$, $\text{Id}$, $+$, $0$, $1$, $U$, $N$. The four types of rules (Formation, Introduction, Elimination, Computation) must now be understood as describing shapes. I see them as doing this by answering the following questions:

- **Formation**: When can we observe $A$?
• **Introduction:** When can we observe a point of $A$?

• **Elimination:** When can we observe points on another shape from $A$?

• **Computation:** What are the symmetries of $A$?

Uniqueness principles (e.g. for $\Pi$-types) can also be seen to answer the same question as Computation rules, and this is how I will regard them.

Now, I will first show how the justification would be given for $\Sigma$-types and then sketch how the same kind of justification could be extended to the rest of the rules, and what difficulties one would face in that regard. In outline, $\Sigma$-shapes are to be understood as “juxtapositions” or “superimpositions” of previously given shapes. We take the four rules in turn.

**$\Sigma$-formation.** The formation rule for $\Sigma$-types is

\[
\Gamma \vdash A \text{ Shape} \quad \Gamma, x: A \vdash B(x) \text{ Shape} \quad \Sigma\text{-Form}
\]

\[
\Gamma \vdash \Sigma_{x: A} B(x) \text{ Shape}
\]

The meaning of what is above the line (ignoring $\Gamma$) is that from the point of view of any point of $A$ we can observe a shape. The meaning of the observation below the line (again ignoring $\Gamma$) is that we can observe a shape from (the act of) juxtaposing each point of $A$ to the shape observable from it. And how do we justify the rule? To borrow Martin-Löf's terminology, this is a rule of immediate inference, namely it is to be understood as correct merely from our intuition about the meaning of the word “juxtaposition”.) The mental act I have in mind here is one where we “twist” a perspective from which two
shapes are distinct in such a way that one becomes viewable from the other. Think of the twist involved here

If one thinks of \( \bigtriangleup \) as the shape observable from the point of view of \( \times \), then what you get below the line is the shape obtained (via twisting) by juxtaposing the two shapes, as indicated by the “\( x \)” in the middle of the shape below the line. And the justification of the rule, I claim, consists in convincing oneself that this (mental) act of “twisting” can always be performed.31

\[ \Sigma \text{-introduction.} \] We have

\[
\frac{\Gamma \vdash A \text{ Shape} \quad \Gamma, x: A \vdash B(x) \text{ Shape}}{\Gamma, x: A, y: B(x) \vdash \langle x, y \rangle: \Sigma B(x) \quad \Sigma\text{-Intro}}
\]

31I have avoided calling such rules of immediate inference analytic, but neither can I call them synthetic lest I find myself defending a view of synthetic a priori propositions. This is a view that I personally find very attractive (and which correlates very well with my experience of doing mathematics), but I hesitate to make this meaning explanation depend upon it. So, instead, what I will say is that the “immediacy” in this inference amount to being able to see straight away that such a move can be made in (Euclidean) perceptual space.
What we have above the line is clear: we can observe a point $a$ of $A$ and a point $b$ of $B[a/x]$ (the shape observable through the point of view of $a$). What is below the line is equally clear: we can observe a point of the shape obtained from juxtaposing $a$ and $b$.

Now it would appear that spatial intuition does not rule in our favour here. For given two points

\[ \begin{array}{c}
\includegraphics[width=0.5\textwidth]{diagram.png}
\end{array} \]

we seem to get also two points in the corresponding juxtaposed shape, viz.

\[ \begin{array}{c}
\includegraphics[width=0.5\textwidth]{diagram2.png}
\end{array} \]

However, the mistake here (and where the picture becomes misleading) is that in juxtaposing $a$ with $B(a)$ we no longer occupy a perspective external to both. Rather, we
occupy the perspective of \( a \) itself. In other words, our point of view (“EYE”) is not situated outside the triangle like this

![Diagram](image1)

but rather \( a \) and our EYE coincide like this

![Diagram](image2)

Understood thus, there really is only one observable point, namely \( b \) (or, better, “\( b \) as seen from \( a \)”). Thus there is a single line-of-sight from \( a \) to \( b \) that we may visualize as follows:
and which we may label \( \langle a, b \rangle \). Clearly, line-of-sight is a new primitive notion, which we have to introduce to make up for the fact that new notation has been introduced in the form of \( \langle -,- \rangle \). My justification of the correctness of \( \Sigma \text{-intro} \) then boils down, once again, to the following simple intuition: if we have a line of sight to a point and point on a shape (the latter of which is observable from the other) then if we occupy the view of the single point we maintain a line of sight to the point on the shape. In other words: two unobstructed points of view allow us to view the object under consideration just the same.

\[ \Sigma \text{-elimination} \]. We have the following rule

\[
\begin{align*}
J_1 \rightarrow [\Gamma, z: \Sigma_{x:A} B(x) \vdash C \text{ Shape}] \\
J_2 \rightarrow [\Gamma, x: A, y: B(x) \vdash d(x,y): C((x,y))] \\
J_3 \rightarrow [\Gamma, z: \Sigma_{x:A} B(x) \vdash \text{split}_d(z) : C(z)] \\
\end{align*}
\]

\[ \Sigma \text{-Elim} \]
where, as indicated, I have labeled the three judgments (inside the square brackets) as $J_1, J_2, J_3$ for ease of reference. Once again ignoring $\Gamma$, we can represent this rule as the following move

\[
\begin{array}{c}
\text{The justification then goes as follows. If we juxtapose } \bullet \text{ with } \begin{picture}(40,40)
\end{picture} \text{ in } J_2 \text{ we get } \begin{picture}(40,40)
\end{picture} \text{ and } J_1 \text{ ensures that the shape that we can see from there is the same and, moreover, a point is observable (in } J_2 \text{): Thus from the point of view of } \begin{picture}(40,40)
\end{picture} \text{ we can observe } \begin{picture}(40,40)
\end{picture} \text{ which is exactly what } J_3 \text{ expresses.}
\end{array}
\]

**$\Sigma$-computation.** We have

\[
\begin{align*}
\Gamma, z &: \Sigma \sum_{x:A} B(x) \vdash C \textbf{ Shape} \\
\Gamma, x &: A, y : B(x) \vdash d(x, y) : C(\langle x, y \rangle) \\
\Gamma, x &: A, y : B(x) \vdash \text{split}_d(\langle x, y \rangle) \equiv d(x, y) : C(\langle x, y \rangle)
\end{align*}
\]

\[\Sigma\text{-Comp}\]
The top part of the line is as before. The bottom part of the line (yet again ignoring \( \Gamma \)) says that from the point of view of \( \bullet \rightarrow \) we can observe that two points are symmetric, i.e. they occupy exactly the same spot on the given shape. In other words, the two points are congruent up to some allowable symmetry. The justification for \( \Sigma \)-comp now runs as follows: in viewing the point labelled as \( \text{split}_d((x,y)) \) we are viewing it through the juxtaposed perspective \( \bullet \rightarrow \) whereas in viewing it through the point labelled \( d(x,y) \) we are viewing it through \( \bullet \rightarrow \). What we observe is the same shape with the same point, except they appear different because we are viewing the from different angles. By shifting the angle accordingly we can observe that the shapes are symmetric (we can make them coincide) and therefore the points congruent. This concludes the justification for \( \Sigma \)-types.

The justification for the rules for \( \Pi \)-types will go along similar lines as the justification given above for the rules for \( \Sigma \)-types, guided by the intuition that we are “gluing” or “amalgamating” shapes (just as for \( \Sigma \)-types we were “juxtaposing”). The key difference between “juxtaposition” and “amalgamation” is that the former involves creating a shape through a union of perspectives without changing the original shapes themselves whereas the latter, as I understand it, involves creating a new shape by fusing our two original shapes together. And indeed, this is reflected in the somewhat unique elimination rule for \( \Pi \)-types which does not really explain what other shapes are observable from a given \( \Pi \)-shape, but rather describes how points of the original shapes can be “recovered” from points of the fused shape.\(^{32}\)

\(^{32}\)For this reason I have sometimes thought it reasonable that \( \Pi \)-types should be considered more fundamental than \( \Sigma, \text{id} \) etc. -types. This is both because of their (inescapable) appearance in the rules for \( W \)-types as well as because of their idiosyncratic Elimination rule. Nevertheless, they do seem amenable to the kind of “spatial” meaning explanation I have here outlined and this places them on equal footing with \( \Sigma \)-types, unless of course one wants to claim that the operation of “amalgamation” is more fundamental than the operation of “juxtaposition” (which I do not).
For + (i.e. coproduct types) we use the idea of “concatenation” and the rules are justified using just this intuition, taking care to note that “concatenation” here is not left-right symmetric, i.e. that it matters which shape is concatenated on the “left” and which one on the “right”, as of course one would expect from the usual behaviour of coproducts which separate points depending on which component they originate from. Similarly, we regard 0 as an empty (or point-less) shape and we regard 1 as a shape that consists of a single point and justify their rules accordingly.

Now, we move to the identity types. The rules for Π, Σ and + considered above can be understood as instructions for how to combine shapes that are viewable one from the other (i.e. how to combine types one of which depends on the other). The rules for identity types, on the other hand, are best understood as instructions for how to draw new things (lines, paths, loops) on pre-existing shapes. This difference is visible even in the usual formation rule for identity types since the premises do not demand a type that is dependent on another type; rather, all it demands is a type and two terms of that type. This suggests, spatially, that identity types are about drawing on a shape rather than combining two (or more) shapes to form another shape. Thus, the basic idea is that identity types between two given points are to be understood as shapes whose points are paths between those two given points – we may call this the “path shape”.

Given the homotopy interpretation of type theory, the notion of a “path shape” should appear as a fairly natural idea by now and the spatial justification of the rules exploit it fully.\textsuperscript{33} The formation and introduction rules for identity types are straightforward to justify spatially: given any two points we can certainly envision a shape formed of paths going between them; and given any point on a shape we can certainly envision a trivial path from that point to itself. The elimination and computation rules are less

\textsuperscript{33}An alternative approach to the justification of the identity types in HoTT, where types are regarded as concepts, is taken by Ladyman and Presnell in [85].
straightforward, as one might expect, since they comprise the key novelty of MLTT. For both rules, the premises are the following:

\[
\Gamma, x: A, y: A, p: \text{Id}_A(x, y) \vdash C \text{ Shape}
\]

\[
\Gamma, z: A \vdash d(z): C(z, z, \text{refl}_z)
\]

Ignoring \(\Gamma\), these premises say that we can observe a shape \(C\) from an arbitrary path and that we can observe a point on each portion of \(C\) observable from the trivial path on an arbitrary point. One thing to note immediately here is that a point visible from a point on a path shape is itself going to be understood as a path. This is reasonable because a “path” for us is not to be understood as an accumulation of points but as an independent notion, indeed as a kind of point in its own right. The visual here is not of a spotlight emanating from a point, casting its light on the surface of a shape, but rather of a beam emanating out of a line casting an elongated shape onto that same surface. It is this visual that we will rely on for the justification of the rules below and it is one that, I think, captures enough of the formal notion of a fibration over a path space without straying too far from pre-formal notions (“beam”, “emanate” etc.).

Now, the conclusion of the elimination rule for identity types is the following:

\[
\Gamma, x: A, y: A, p: \text{Id}_A(x, y) \vdash J_{z,d}(x, y, p): C(x, y, p)
\]

Once again ignoring \(\Gamma\), this says that we can observe a point \(J_{z,d}(x, y, p)\) on the portion of \(C\) that is observable from an arbitrary path \(p\) on \(A\) from \(x\) to \(y\). Given what was said above about points observable from paths, \(J_{z,d}(x, y, p)\) is then best thought of also as a kind of path, i.e. a “stretched” point. The situation is illustrated by Figure 3, in which the red line is the “stretched” point obtained through the observations of \(d(x)\) and \(d(y)\).
the dotted lines indicate the portion on \( C \) observable from the reflexivity path on the point \( x \) of \( A \), the black line is the path \( p \) and the trace of circles indicates the portion of \( C \) visible from \( p \). The fact that this stretched point can be observed given the two “endpoints” is then justified as follows: since the existence of the path \( p \) guarantees that there is an unbroken beam visible from \( d(x) \) to \( d(y) \) there is certainly a way to connect \( d(x) \) and \( d(y) \) through this unbroken beam. In terms of the picture above: since there is a “tunnel” from \( d(x) \) to \( d(y) \) there is a way to go through the tunnel from \( d(x) \) to \( d(y) \). This “way of going through the tunnel” – which intuition tells us ought to exist – we call \( J_{z,d}(x,y,p) \).

**Figure 3.** The black line shows a path from \( x \) to \( y \) (in some shape). The “tunnel” shows the shape observable from this path (on some other shape). And the red line shows the “stretched point” we can observe on the “tunnel shape”, which is what the elimination rule states ought to exist.

On the other hand, the conclusion of the computation rule is the following:

\[
\Gamma, x: A \vdash J_{z,d}(x,x,\text{refl}_x) = d(x): C(x,x,\text{refl}_x)
\]

Once again ignoring \( \Gamma \), this says that “stretching” a point along the trivial path gives us the same point. Spatially, this is a triviality: if I can observe point from a certain vantage
point and then stay put in that same vantage point, then I am still observing the same point. This concludes the spatial justification of the identity types.

Finally, we are left with the justification for the rules for $\mathbb{N}$, $\mathcal{U}$ and the univalence axiom for $\mathcal{U}$. Here, the spatial meaning explanation faces serious difficulties. I take the three groups of rules in turn. The difficulty with $\mathbb{N}$ is, roughly, that we need to sneak in some kind of “temporal” element. This temporal element has to do with the need to talk about an “indeﬁnitely extensible” process in order to assert the existence of arbitrarily many points in $\mathbb{N}$. Formally, this comes up in the introduction rule for $\mathbb{N}$ that involves the successor:

$$
\Gamma \vdash n : \mathbb{N} \quad \text{N-Intro-2} \quad \Gamma \vdash s(n) : \mathbb{N}
$$

Spatially, the rule N-Intro-2 says that if we can observe a point $n$ on $\mathbb{N}$ then we can observe another point $s(n)$. In other words, we need to somehow be able to say that the shape $\mathbb{N}$ is “indeﬁnitely extendible”. Of course, indeﬁnite extensibility can be cashed out geometrically in terms of the indeﬁnite divisibility of some kind of continuum. But such an approach seems hopeless in our setting, since it requires us to assume an intuition of the continuum in order to describe a discrete shape such as $\mathbb{N}$. This might be a bullet worth biting, but only if other options have been exhausted. Therefore, I am currently not sure how to describe an inﬁnite discrete shape (such as $\mathbb{N}$ is supposed to be) without

---

34Essentially, the task of making this precise amounts to giving an interpretation of PA/HA into theory with pure geometric content. In the setting of ﬁrst-order logic, this is not entirely without precedent. See for example the work of Hellman and Shapiro [64]. More recently, in private communication, Harvey Friedman has suggested a way of interpreting $Z_2$ (roughly, second-order arithmetic) into a ﬁrst-order theory built out of purely geometric notions. Such efforts, especially the ones of Friedman that were motivated by similar considerations as the ones that motivate me here, could certainly help in ﬂeshing out a purely geometric understanding of PA/HA, one that could then be used to give a more purely spatial justification to rules such as N-Intro-2.
either cashing out indefinite extensibility in terms of some temporal notion of progression, or relying on some ambient continuum which is indefinitely divisible. It is an interesting dilemma, and one that merits further exploration.

Now, on to universes. A universe must be understood as some kind of “shape of shapes”. Intuitively, this seems more difficult to conceptualize than a “collection of collections” which is how a universe in traditional MLTT can be understood. I will offer here a few suggestions as to how to understand the rules for the universe $\mathcal{U}$ as describing such a shape of shapes. The justification of the rules expressing the closure of the universe under specific type constructors can be justified just as before, so I will say no more about that. The other two rules are the following:

\[
\begin{align*}
\vdash \mathcal{U} & \text{ Shape} \\
x : \mathcal{U} & \vdash \text{El}(x) \text{ Shape}
\end{align*}
\]

The rule $\text{Uni-Form}$ is to be understood as asserting that a shape of shapes is observable from the “neutral” viewpoint, which is to say from every viewpoint. Whether or not this can be justified intuitively I do not know – but what seems to me right is that it can as much (or as little) be justified as the idea of a collection of all small sets (that is not itself a set). The visual that I find helpful here is that of a fractal-like accumulation of points – and the closer one “zooms into” one of these points, the more one sees them as shapes in their own right. Of course, this requires that a “point” is no longer necessarily understood as an irreducible part of a shape, but as a shape in its own right. And indeed, this is exactly what the dependent shape El expresses in $\text{Uni-El}$ and this is also the reason why it is necessary to introduce it, viz. in order to be able to “blow up” points into
shapes. However, as far as I am concerned, the problem with this talk of “fractal-like” self-similarity and of points morphing into shapes is that these concepts stray too far from what might be considered “intuitive”, at least in the sense in which I have so far been using the term. I am currently not sure if a more intuitive visualization can be found.

Finally, we have the univalence axiom, which in 1UHoTT takes the following form:

\[
\begin{align*}
\Gamma & \vdash A : U \\
\Gamma & \vdash B : U \\
\Gamma & \vdash u_{A,B} : \text{isequiv}(\text{idtoequiv}_{A,B})
\end{align*}
\]

where

\[
\text{isequiv}(f) =_{df} \Pi_{x : A} \text{iscontr}(\text{fib}(f, x))
\]

and for any type \(A\) we define

\[
\text{iscontr}(A) =_{df} \sum_{a : A} \Pi \text{Id}_{A}(x, a)
\]

and for \(f : A \rightarrow B\) and \(y : B\) we define

\[
\text{fib}(f, x) =_{df} \sum_{x : A} \text{Id}_{B}(f(x), y)
\]

Also, setting

\[
A \simeq B \equiv \sum_{f : A \rightarrow B} \text{isequiv}(f)
\]

we have that

\[
\text{idtoequiv}_{A,B} : \text{Id}_{U}(A, B) \rightarrow A \simeq B
\]
is the canonical term obtained by induction on identity by a standard argument. Given that univalence is introduced as an axiom, relying on the above auxiliary definitions of “contractibility”, “fibers” and “equivalence”, there is little hope that the spatial meaning explanation I have been outlining could be used to justify it directly. The problem lies with the definition of the auxiliary notions above. To illustrate, consider the notion of contractibility of a shape, as formalized by \texttt{iscontr}. The intended meaning can be straightforwardly gleaned if we unpack \texttt{iscontr}(A) in the PAT style: \texttt{iscontr}(A) says that there exists a point \(a\) of \(A\) such that any other point \(x\) can be connected to \(a\) by a path. This is perfectly reasonable geometrically, needless to say; but it is only reasonable geometrically if we read the type constructors \textit{logically}. This is more generally the issue with introducing any kind of axiom to type theory. The only reasonable way forward on this issue is to define a HoTT in which axioms such as univalence are derived as consequences of rules introduced in (as much as possible) the usual pattern of formation, introduction, elimination, and computation rules. The ability to do this for univalence (the so-called “constructive interpretation of univalence”) was indeed the major driving force behind the development of CTT, about which I will have a few more words to say below.

3.3.5. Discussion and Objections. So this, in outline, is how the geometric meaning explanation would proceed for the usual type-formers. Let me now consider some objections and draw some conclusions. As regards objections, I see three main worries.\[\text{35}\]

First, we can regard \(\simeq\) as a dependent type over any identity type in \(\mathcal{U}\) by noting that

\[
A: \mathcal{U}, B: \mathcal{U}, p: \text{Id}_\mathcal{U}(A, B) \vdash A \simeq B \; \text{Shape}
\]

where the path \(p\) in the context does not appear on the RHS. Now, note that the identity map \(1_A \equiv \lambda x: A.x\) is an equivalence for any \(A: \mathcal{U}\), which means that we can always find an inhabitant \(d(A)\) of \(A \simeq A\). But then by \texttt{Id-Elim}, we can derive the judgment

\[
A: \mathcal{U}, B: \mathcal{U}, p: \text{Id}_\mathcal{U}(A, B) \vdash J_d(A, B, p): A \simeq B
\]

By \texttt{Π-Intro} we then get the desired section

\[
\text{idtoequiv} \equiv \lambda (A, B). \lambda p. J_d(A, B, p): \Pi_{A, B: \mathcal{U}} \text{Id}_\mathcal{U}(A, B) \rightarrow A \simeq B
\]
First, one might object that the basic notions of the spatial meaning explanation ("observation", "shape", "symmetry" etc.) are not fundamental enough to serve the purpose of a pre-formal meaning explanation. There is not much to say to this other than that they feel simple enough to me. As mentioned in the beginning, at some point with genuine meaning explanations, the reader will have to rely on their own cognitive faculties to convince him/herself that the notions being discussed are meaningful. To my mind the four notions I have described certainly appear to me to be simple enough. I certainly feel, when meditating on the justification of the rules, that my "pure intuition" (whatever that is) is adequate to the task of visualizing them. And I think I'll leave it at that.

Second, one might object that the basic intuitive notions I have used do not tack on well with the formal notions in the set-theoretic models of HoTT. For example, one might protest that the formal notion of a "fibration" does not match my intuitive notion of "observable from an arbitrary point" or that the notion of a "total space" does not match the notion of "juxtaposing shapes". To this I can only reiterate my conviction that these intuitive notions really do tack on to the formally specified ones from algebraic topology. After all, I think it is quite clear that the formal notions originating from algebraic topology, and developed within set theory, are guided by very similar intuitions.

Thirdly, one might in general recoil at the prospect of using what appears to be basic Euclidean intuition to justify logical rules. Surely, if progress on the foundations of mathematics has taught us anything in the last century, it is that we should not rely on geometric intuition as a standard of rigour – and I certainly seem to be relying on Euclidean intuition! For example, it seems highly objectionable that the meaning explanation seems to rely on fixed intuitions about the continuum, e.g. about how shapes can be moved through it. This seems to put the cart before the horse: homotopy theory is supposed to give us the essence of shapes in the continuum by simplifying them to their
essential features; it seems pointless to justify these features in terms of the very notions
they are meant to simplify. To this, I must once again assert a distinction between
human intuitions of the continuum and the way these intuitions have been formalized
in (set-theoretic) topology. Of course, the formalization of the continuum was meant
exactly to save us from inaccuracies born of an excessive reliance on intuition – but this
does not mean that such intuition does not exist, nor that it cannot be relied upon for
certain things. Yes, I am relying on our intuitions of the continuum and I am also of the
opinion that these intuitions are more-or-less Euclidean (I consider it an entirely separate
question whether this has anything to say about the very nature of space). But once
again, with respect to our purposes here, the proof of the pudding really is in the eating:
whatever kind of intuition it is that I am relying on (and I am certainly relying on some
such intuition) what matters is that the reader should be able to employ it to justify
the rules we have been considering – for that would ensure that it is a universal human
intuition, thus guaranteeing some kind of intersubjective validity for the rules (which is
in any case the most we can reasonably hope for).

Another big question is the suitability of 1UHoTT as a formal theory for UF amenable
to an intuitive justification. As we saw, the justification of \( \mathbb{N} \), \( U \) and the univalence
axiom runs into serious difficulties. Indeed, I do not believe that 1UHoTT will ever be
fully amenable to the kind of spatial meaning explanation sketched below. The reason,
alluded to above, is that identity types in MLTT remain inescapably “logical”. Among
other things, this forces us to state univalence as an axiom, instead of proving it as a
consequence of properties of other type constructors. The HoTT that is perhaps better-
suited to receive my meaning explanation will be something closer to Cubical Type Theory
(CTT) where everything other than \( \Pi \) and \( \Sigma \) types are introduced with an explicit spatial
intuition, including the path types which are distinct from the identity types of MLTT.\(^{36}\) As such, in CTT the below-sketched perspective of rules-as-instructions-to-construct-shapes is “written in”, rather than tacked on \textit{a posteriori}. In other words, the meaning explanation I have in mind is (implicitly) used in designing the rules of CTT, except of course the intuition used there is that of homotopy theory understood in terms of cubical \textit{sets}, which already is guided by set theory. As such, I expect that only a HoTT much like CTT, whose design is spatially motivated from the very beginning, will ever be fully amenable to a meaning explanation along the lines suggested here. That said, there are two main issues with CTT, as I currently see it: firstly, it fares no better with respect to \(\mathbb{N}\) and universes; and secondly, the basic syntax of CTT is far more complicated than that of traditional MLTT, since, for example, many new operations are added in defining contexts. But a spatial meaning explanation for CTT is clearly an avenue worth exploring.

However, one thing which the spatial meaning explanation \textit{does} achieve, as far as I am concerned, is independence from set theory and first-order logic at an intuitive level. A simple inspection of what I have said so far should suffice to convince the reader that at no point have I employed talk of properties, collections etc. – namely, at no point have I appealed to some intuitive theory of sets. Thus, insofar as the spatial meaning explanation can be extended to other, stronger HoTTs, we have good reason to claim that UF can achieve \textit{complete} autonomy from set theory, even at the level of intuitive justifications. This would be the first step towards making a convincing case that UF can be a self-standing foundation of mathematics that could plausibly serve as a benchmark of consistency on its very own.\(^{37}\)

\(^{36}\)Nevertheless, the identity types of standard MLTT (with a judgmental computation rule) are expected to be interpretable into CTT, as described in Section 9.1 of [21].

\(^{37}\)For reasons outlined in Section 3.2, I have some reservations that this can ever be achieved.
Finally, let me make what I think is a fundamental observation, indeed quite possibly the most important observation in this entire work: the symbolic presentation of the rules of 1UHoTT ceases to be necessary for a meaning explanation of HoTT. This idea, I suspect, is likely to provoke extreme skepticism at this moment in time and the few sentences that I will spare it can hardly put the tiniest dent into more than a century’s worth of momentum (perhaps also prejudice but without a doubt a body of severely calcified intuitions) that has shaped our idea of a “formal system” as something comprised of symbols that are in no way directly correlated to what we want to take them to represent, i.e. of the idea that in defining a formal system, the symbolic substitution has already occurred, in fact that it occurs as comprising the essential condition for the definition of a formal system. But here, with HoTT, along the lines of the meaning explanation that I suggest, we can allow ourselves to envision a formal system whose very “symbols” are the shapes that one intends to describe with it. In other words, we can envision a completely native graphical syntax for HoTT, in which the inference rules do not move us from a (or several) lines of symbols (separated, say, by turnstiles) to other such lines of symbols, but rather indicate what kind of new shapes we can draw from old. We replace judgments themselves with shapes, altogether ignoring the fact that it is certain letters in those judgments (e.g. “A” in “Γ ⊢ A: Type”) that are supposed to be stand-ins for shapes. This idea suggests a new form of reasoning, spatial in character, and altogether independent from first-order logic. But in its “purest” form it is likely to appear alienating, if not some crazed idiocy. Thus, more constructively, I have instead defined an intermediate form of this logic, one whose syntax lies half-way between geometry and “pure” symbolism – the definition of this logic occupies all of Chapter 5.

For now, I require only a bit of patience, perhaps some minor suspension of disbelief, that such a “logic” of drawing shapes (instead of manipulating concepts, predicates and
relations) is indeed possible and that it can be used to provide the justification for the rules of HoTT as has been sketched in this section. Such a possibility allows us to answer – I think decisively – a very powerful objection that has been raised against categorical foundations of mathematics (indeed against all non-set-theoretic foundations). To this, the next Section is devoted.

3.4. The Problem of the Home Address

Hellman’s \[62\] ingenious objection to categorical foundations of mathematics essentially boils down to this: any foundation will have to provide some kind of notion of satisfaction, namely the idea that a piece of syntax $\phi$ (e.g. a well-formed formula in first-order logic) is satisfied by some “piece” of semantics (e.g. a set $M$ that interprets all the constituents of $\phi$ in the usual Tarskian manner). But the only way we have – Hellman argues – of expressing such a notion of satisfaction is by talking about a collection of stuff that stands in a certain relation to a particular piece of syntax. Since any kind of foundation will have to assert – on some level – that certain pieces of syntax (axioms) are satisfied, categorical foundations of mathematics are not self-sufficient: they will always depend on some ambient set-like theory of collections if they are ever to contain a usable notion of satisfaction, along the lines of Tarski’s definition.

Relatedly, Stewart Shapiro in \[162\] has raised very similar worries about category theory while considering its role as a metamathematical tool. Shapiro’s challenge to category theory as a framework for the foundations of mathematics takes the form of a dilemma. On the one hand, category theory cannot support a structuralism that goes “all the way down” and maintain a foundational status, since even if we accept that all of mathematics takes place in axiomatically defined algebraic structures, in order to have criteria of acceptability for such theories we still require a metamathematical background that is “assertory”; on the other hand, if we don’t want such an assertory
background we are forced to chuck the foundational ladder altogether, in which case we accept that metamathematics belongs to “philosophy.” Shapiro’s requirement for an “assertory” background I take to be exactly analogous to Hellman’s (disguised) demand for a uniquely category-theoretic notion of satisfaction. In other words, Shapiro’s dilemma for category theory says that category theory must either come up with a non-set-theoretic way of understanding what it means to assert that some sentence is satisfied by some structure, or else to bite the bullet and concede that category theory can only provide a foundation of mathematics without formal metamathematics.

This challenge, raised by both Hellman and Shapiro, I will call the problem of the home address, following [62]. For the sake of precision, let me state it, combining the formulations of Hellman and Shapiro into one:

(PHA) A categorical foundation for mathematics cannot be independent of set theory unless it is able to assert the existence of a certain structures without relying on a set-theoretic notion of satisfaction. In particular, categorical foundations of mathematics should be able to carry out their own metamathematical investigations without relying on the standard model-theoretic tools developed inside set theory.

There are a few standard responses to the PHA that have been considered in the literature. McLarty [132] responds by pointing out that categorical foundations like ETCS and CCAF are not to be compared to the axioms for a general category. The former do postulate existence, whereas the latter do not – indeed, as McLarty points out, it seems odd to say that there exists a “generic” category, i.e. a “generic” model of the first-order theory of categories. But I think that McLarty misreads the PHA in that the issue is not about categorical assertions being able to assert the existence of some universe of objects (e.g. the category of sets), but about the way in which this assertion can be
formalized. If one wants to study ETCS metamathematically, one is forced to study it as a first-order theory – and the only way that is available to us to study first-order theories is provided by a set-theoretic, not a category-theoretic, foundation.

Picking up on this point, Landry [89], responding to Shapiro, argues that categorical structures can be constructed internally to other categorical structures and so category theory is perfectly well-able to provide a categorical notion of satisfaction to be used for metamathematical purposes. We can interpret, say, a topos inside another topos without mentioning a universe of sets at all. But this response misses the deeper point of the PHA: even when one interprets objects internal to a category one is interpreting a piece of syntax with a piece of semantics by taking the piece of syntax to be represented by collections of things inside your piece of semantics. As such, one is still using a set-theoretic notion of satisfaction – what obscures this fact is that the “piece of semantics” has a priori been given the structure of a category.

On the other hand, one might object that the PHA seems to be saying no more than Skolem’s well-known “pre-existing domain” objection to set theory, namely that it seems impossible to talk about models of set theory without having a notion of a set already in place. And so category theory is no more susceptible to the PHA than set theory is susceptible to a Mayberry-style objection that the model theory of first-order set theories (e.g. ZFC) requires some pre-existing notion of collection, thus pulling the rug from under our feet. But even if one finds such objections compelling, the PHA still stands: for whether or not one is moved by Skolem’s objection one must still admit that he is left equally moved by it whether it is, say, ETCS or ZFC under consideration. And this, I take it, is all that the PHA is pointing out: that categorical foundations of mathematics

---

38 We can also interpret the cumulative hierarchy inside a topos, cf. e.g. [7, 75].
bring nothing more to the table with respect to how we answer the questions “Does a model $M$ satisfy a theory $T$?”

None of these responses, I conclude, succeed in blunting the problem of the home address. In UF, however, a decisive response becomes available. This is because UF comes with a notion of satisfaction that is uniquely category-theoretic in nature and wholly distinct from Tarski’s. The pre-formal justification given in Section 3.3 already contains the seed of the idea: the satisfaction of a particular piece of syntax is to be understood as the construction of a particular shape, and not as a “truthmaking structure”. In what follows, I will make this intuition precise and explain how it answers the PHA.

Firstly, it is instructive to examine the difference in the notions of satisfaction between first-order logic and MLTT. To see this difference let us unpack what it means for a “proposition” $P$ to be “provable” in MLTT. I assume here that $P$ is some kind of sensible pre-formal proposition (e.g. “There is a least odd prime”) that can be formalized in a sufficiently expressive logical calculus and I will use the same letter $P$ to refer both to the formalization of this proposition as well as to its informal counterpart. Under the propositions-as-types philosophy of MLTT, $P$ is to be regarded as a type and we say that $P$ is provable in MLTT if we can produce a term $p$ of $P$, i.e. if the judgment

$$\Gamma \vdash p : P$$

I owe this comparison to Skolem’s objection to John Burgess.

It is worth mentioning that the affinity between categorical thinking and constructive thinking (in the broad sense) has been apparent almost since the beginning of category theory. Essentially, this boils down to the fact that most of the things one wants to do in the setting of category theory are universal constructions that define objects only up to isomorphism, thus bypassing thorny questions pertaining to the nature of these objects. This inherently constructive side to category theory has been studied by McLarty [133] in which he offers six very interesting hypotheses as possible explanations. Also related is Moerdijk’s [137] exploration of the same issue and the connections he draws with synthetic geometry.
is derivable in MLTT. This judgement already contains a lot more information than just the proposition $P$ itself – in particular it contains information about a specific proof $p$ of that proposition.\footnote{What I have said in this paragraph applies to most variants of dependent type theory. Clearly, I am temporarily sweeping under the rug well-known difficulties about what is to be regarded as a “proposition” in HoTT. Such complications, though serious, do not affect the points being made here.}

By contrast, the only relevant information encoded by a first-order sentence $\phi$ is what $\phi$ says (or expresses). Therefore, whereas a judgement of the form $\Gamma \vdash p: P$ in MLTT is essentially a metamathematical statement about a proposition (viz. the statement that a certain proposition $P$ has a proof $p$ in context $\Gamma$), the satisfaction of a formula $\phi$ in $\text{FOL}_=$ is a mathematical statement about some mathematical state of affairs e.g. that any even number is the sum of two prime numbers.\footnote{I am here merely elaborating on the distinction between propositions and judgments about propositions. Martin-Löf \cite{126} himself traces this distinction back to Kant.} In short, MLTT is a formal system that manipulates metamathematical judgements some of which are meant to be understood as statements that certain propositions (understood as types) have proofs (understood as terms of those types). On the other hand, $\text{FOL}_=$ (together with, say, a binary relation $\in$ and the axioms of $\text{ZFC}$) is a formal system that manipulates mathematical statements: the basic syntax does not include statements meant to be understood as saying that a certain formula has a proof.\footnote{Of course by the usual coding techniques notions such as “is a derivation” and “is provable” etc. can be expressed as formulas in these first-order formal systems. But my point is that these formal systems are not designed specifically in order to be able to do that. Peano Arithmetic can certainly express what it means for $p$ to be a (code for a) proof of $P$ – but the axioms of Peano Arithmetic are not meant to be understood as metamathematical statements about mathematical propositions.}

What emerges from these observations is the following contrast. To say that a proposition $P$ is satisfied in MLTT is to say that the sequent $\Gamma \vdash p: P$ is derivable in MLTT. On the other hand, to say that a proposition $P$ is satisfied in first-order logic by some structure $\mathcal{M}$ is to say that the interpretation of $P$ in $\mathcal{M}$ comes out true. In other words, sentences in first-order logic describe a state of affairs and their satisfaction says that the
state-of-affairs they describe obtains. But in the MLTT framework to say that the sequent $\Gamma \vdash p: P$ is satisfied is to say that a proof of that proposition has been constructed. To be sure, $\Gamma \vdash p: P$ also describes a particular state-of-affairs but it does not describe the state-of-affairs that $P$ describes: it describes, instead, the state-of-affairs of having a proof of $P$. This is the essential distinction between the syntax of a “proof-relevant” system like MLTT and a “proof-irrelevant” system like first-order logic.\footnote{This difference is expressed nicely by Sundholm in \cite{39}, when discussing Martin-Löf’s general philosophical outlook towards the foundations of mathematics: [MLTT] is an interpreted formal language. [...] In FOL the metamathematical “expressions” employed in mathematical logic are mere objects of study, but do not express. On the contrary, they are objects that may serve as referents of real expressions. In [MLTT], on the other hand, the expressions used are real expressions that carry meaning. In a nutshell, the language is endowed with meaning by turning the proof-theoretic reductions into steps of meaning explanation. Just like Frege’s ideography, or of the language of Principia Mathematica, the type-theoretic formulae are actually intended to say something. They do not essentially serve as objects of metamathematical study. (\cite{39}, p. xx)}

As a result of this syntactic difference, the semantics of a proof-relevant system imitate the structure of the deductive system itself, whereas those of a proof-irrelevant system provide truth-conditions for the propositions that the deductive system handles.\footnote{This difference, it must be said, is a difference in attitude. It does not necessarily underlie some rigid technical or metaphysical distinction. For the semantics of proof-irrelevant systems can also be understood as imitating the structure of the deductive system (by noting that the inductive definition of truth for combinations of formulas mirrors the introduction and elimination rules for logical connectives) and those of proof-relevant systems can be seen as providing truth-conditions for propositions (if by “propositions” we understand whole judgements like $\Gamma \vdash p: P$). But my point is that they are not meant to be understood in this manner, especially when formalizing mathematics.} Proof-relevant systems like MLTT thus come with their own special sort of semantics. These semantics are taken in highly structured categories usually called contextual categories.\footnote{See \cite{28, 168} for an introduction to those and see \cite{77} for a definition of the necessary structure on a contextual category needed for UF. Other names used for these categorical semantics include: comprehension categories \cite{71}, categories with attributes \cite{67}, categories with families \cite{37}, and C-systems \cite{179}.} What it is for the judgement $\Gamma \vdash p: P$ to be satisfied in a contextual category $\mathcal{C}$ (with appropriate extra structure) is for there to be a section (right inverse) to the canonical
projection

\[ \pi_P : (\Gamma, P) \to \Gamma \]

Whether or not one is familiar with the above condition, it is clear that the notion of satisfaction for the judgement \( \Gamma \vdash p : P \) is not at all the same kind of “satisfaction” that we find in the Tarskian paradigm but a new and independent notion.

In what relevant way is it different? And, more pertinently, in what way is it importantly different from the way satisfaction is cashed out in standard categorical logic? For it would appear, initially, that the only difference between the way that satisfaction of a judgement in MLTT is interpreted in contextual categories and the way in which satisfaction of a sentence in first-order logic is interpreted in appropriately structured categories (e.g. coherent categories in the case of the coherent fragment of first-order logic) is that in the former we require the existence of a section to a not-necessarily-monomorphic arrow whereas in the latter we require a section to a monomorphism. To be more precise, if \( \phi \) is a sentence in first-order logic over some signature \( \mathcal{L} \) and \( \mathcal{M} \) any \( \mathcal{L} \)-structure in an appropriately structured category \( S \) then the truth of \( \phi \) in \( \mathcal{M} \) is equivalent to the existence of a map

\[ f : 1 \to \phi^\mathcal{M} \]

in \( C \). This \( f \) can be understood as a section of the monomorphism \( \phi^\mathcal{M} \to 1 \) given by the interpretation of \( \mathcal{L} \) in \( \mathcal{M} \). On the other hand, as described above, if \( P \) is a type in the empty context in MLTT (regarded as a sentence) then the interpretation of the judgment \( \emptyset \vdash s: P \) in a contextual category \( C \) is a section

\[ s^C : \emptyset^C \to (\emptyset, P)^C \]
of the canonical projection given by the interpretation of $P$ in $C$. Clearly both conditions appear exactly analogous. In both cases the truth of a sentence is parsed in terms of the existence of a certain arrow in a category. In the former case this arrow is unique (registering provability) whereas in the latter case it is not necessarily unique (registering only one out of many possible proofs). So what is it about the notion of satisfaction in MLTT that sets it apart from the notion of satisfaction in categorical logic that allows the former, but not the latter, to avoid the PHA? After all, they both seem to boil down to requiring the existence of certain arrows.

The above way of comparing the two notions already reveals the crucial disanalogy: $f$ is not required to be the interpretation of a particular piece of syntax already in our possession, whereas $s^C$ is required to be the interpretation of the specific term $s$ that we are given in the judgment we are interpreting. To make the point clearer: the mere existence of a map $1 \to \phi^M$ in $S$ guarantees the truth of $\phi$ in $M$, but the mere existence of a section $t: \emptyset^C \to (\emptyset, P)^C$ in $C$ does not guarantee the truth of $P$ (more precisely: the truth of the judgment $\emptyset \vdash s: P$).\(^{47}\) Thus, in categorical logic the truth of sentences depends on features of the ambient category (into which the syntax is interpreted) that may not appear in the syntax at all whereas in contextual categories, the truth of sentences depends only on features of the ambient category that already appear in the syntax. Therefore, insofar as contextual categories for MLTT even deserve the name “semantics”, they do so only by virtue of the fact that they provide a structure which “simulates” every move that can be made in the syntax. This means that the categorical semantics of MLTT can best be understood as just another way of packaging the syntax of MLTT and indeed it would even be possible to define the syntax of MLTT to be a contextual category.\(^{48}\)

\(^{47}\)Indeed, it is perfectly possible for there to be sections of projections in contextual categories that are not the interpretations of terms in the syntax; as a result, it is possible for there to be uninhabited types in MLTT whose canonical projections have a section.

\(^{48}\)Voevodsky’s recent work [179] on the “Initiality Conjecture” takes such a perspective of the syntax of MLTT.
An analogous move makes little sense in (categorical or set-theoretic) first-order logic: set theory or elementary toposes cannot be understood simply as an alternative way to re-package the syntax of first-order logic. For a first-order sentence to be satisfied in an elementary topos is for that elementary topos to possess a certain property (the existence of a certain map). And this state of affairs has no direct counterpart in the syntax.

So far so good. We seem to have identified a substantial difference between notions of satisfaction in dependent type theories and those in categorical first-order logic. Are we out of the woods then? Not yet. For one might now object that what is really being satisfied in a contextual category is not a type (once again, for simplicity, regarded as a proposition) but a judgment that a certain type is inhabited. Thus, if we regard the interpretation of MLTT into contextual categories as the interpretation of judgments, then there seems to be no difference at all between first-order logic and type theory with regards to satisfaction. In other words, if we regard the whole of a judgment \( \Gamma \vdash s : P \) as analogous to a sentence \( \phi \) in first-order logic, then the satisfaction of the former depends just as much on an “unsyntactical” property as the latter. Indeed, this will more or less be true of any reasonable deductive system. And since any such system can be thought of as a(n essentially algebraic) first-order theory, the semantics of a deductive system can always be understood simply as the models of this associated first-order theory. This leads us to an even stronger form of the PHA: any formal system can itself be understood as a first-order theory (if we regard it as a “calculus of symbols”) and to talk about the

\[ \text{Of course, if we regard, say, an elementary topos as an intuitionistic type theory in its own right (in the manner of \([87]\)) rather than as a structure into which plain first-order logic can be interpreted, then clearly there is a very real sense in which an elementary topos does become an alternative way to package the syntax of intuitionistic type theory. The issue here is what to regard as a translation of one formal system into another (a formal interpretation) and what to regard as a semantics. As explained at the beginning of Section 3.3 any kind of formal semantics is best understood as a formal interpretation of one language into another. But an intuitive semantics requires something more, namely an explication of the language in terms of pre-formal notions. And the problem of the home address, in its strongest form, is about an intuitive notion of satisfaction, not about the availability of truth predicates in alternative formal systems.} \]

\[ \text{Insofar as it manipulates recursive sets of symbols and contains a recursive collection of rules that depend on decidable side-conditions.} \]
semantics of any first-order theory we require some kind of set theory; thus, insofar as we
are using any kind of formal system, we are in the framework of a first-order theory and
therefore we require some ambient set-theoretic foundation to be in place in order to be
able to say that this formal system “has a model”.

It is this form of the PHA that we must now tackle. Before doing so, let me summarize
what has been said so far as a way of clarifying the dialectic. Firstly, I have explained the
difference between satisfaction in dependent type theories and satisfaction in first-order
logic. As we saw, this difference between the two seemed less pronounced for categorical
first-order logic since in that case both notions of satisfaction seemed to amount to the
same thing, namely the existence of certain arrows. I then argued that there is still
an essential difference between the two notions by explaining that the existence of the
appropriate arrow in the dependent-type-theoretic setting is required to be the semantic
counterpart of a term in the syntax, whereas the same does not necessarily hold true of
the appropriate arrows in categorical first-order logic. However, the two conditions do
coincide insofar as we understand the dependently-typed calculus of terms as a first-order
theory (e.g. as the first-order essentially algebraic theory of C-systems) and the PHA
re-arises in an even stronger form. In order to answer this even stronger form, we need
to separate not just dependent type theories from first-order logic, but also dependent
type theories in general from homotopy type theories. So what I will argue for now
is that the term calculus of HoTTs is not best understood as a first-order theory. In
particular, there is a way for us to assert “HoTT has a model” without going through
the usual process of assigning “truth values” to strings of symbols and then ensuring
that these “truth values” are preserved under certain operations; or, equivalently, as in
Voevodsky’s project in [176, 179], by constructing a term model and a mathematical model and establishing the existence of a homomorphism from one to the other.\textsuperscript{51}

The crux of my argument is based on the perspective outlined in Section 3.3, namely that a HoTT, as a deductive system, behaves less like a first-order theory and more like a geometric object in its own right. This becomes especially clear when we move away from MLTT towards a more “natively” homotopic formalism like Cubical Type Theory (CTT). What one has in the beginning is some raw syntax together with certain specifications (judgmental equalities) concerning which symbols are to be assumed equal (or “convertible”). At this level, we are able to separate symbols that we want to be identical, i.e. we are able to write a judgement “\(\Gamma \vdash a = b : A\)” even though \(a\) and \(b\) are not identical entities. This is entirely unproblematic, of course, because \(a\) and \(b\) exist at the “symbol” level. We regard them as uninterpreted symbols and the only fact about them that interests us is that we can assert that they are distinct (as symbols).\textsuperscript{52} The step which now seems unavoidable is that any kind of interpretation (anything we should wish to call an interpretation) must convert these asserted equalities into real equalities, where for the time being what is “real” can be understood relative to some other ambient formal system.

What is required in order to avoid this “unavoidable” step is a notion of equality that is straightforward enough to allow us to operate at the “symbolic level” (i.e. when we want to regard “\(\Gamma \vdash a = b : A\)” as a string of symbols) and at the same time sufficiently meaningful so as not to require us to interpret it in terms of another notion. The notion of symmetry/congruence of shapes, I submit, is exactly such a notion. In order to use symmetry at the “symbolic level”, the key step is the one alluded to at the end of Section

\textsuperscript{51}In Voevodsky’s approach there is also a hidden step. To construct a term model one must also prove that the term model is the “implementation” of the raw syntax of the particular formal system under consideration (cf. [176], p. 4).

\textsuperscript{52}If we are working in some formal system when reasoning thus what I am saying amounts to the requirement that our symbols are the elements of a set with decidable equality.
3.3, viz. we have to understand the symbols themselves as shapes. In particular, whatever it is that we mean by “symbol” it ought to be something about which it makes sense to say that one symbol is congruent/symmetric to another.

If one brings to mind what we usually call a “symbol” – letters like “a”, “b” etc. – this suggestion must appear very difficult to understand, if not absurd. And yet the seed of the idea is already present even in this way of understanding symbols. For we can easily describe the non-identity of \(a\) and \(b\) in terms of spatial properties of the symbols themselves, e.g. that \(b\) contains a straight line segment but \(a\) does not; and that therefore, assuming a certain rigidity in the ambient space, \(b\) cannot be made to coincide with \(a\) no matter how we move the two around. Indeed, the very utility of \(a\) and \(b\) as letters of the alphabet comes from the fact that it is easy for us to distinguish them as shapes. What I urge, therefore, is simply to take this basic idea very seriously.

This allows us to take the following perspective on a deductive system: it consists of shapes and instructions about how to construct new shapes from old ones (e.g. by juxtaposing, amalgamating, drawing paths) together with constraints about what kind of observations it should be possible to make given certain observations already made. To break this down with a little more specificity: formation rules assert what kind of shapes one can observe given other observations; introduction rules allow us to draw points on observable shapes; elimination rules allow us to draw points on shapes viewed through other shapes; computation rules assert that certain symmetries should be observable. More relevant to our concerns, this point of view allows a seamless integration of the syntax with the semantics. The syntax of the system from this point of view can be thought of as the “minimal” collection of shapes that is able to articulate the full gamut of observations that we are interested in. Crucially, the semantics of the system is then not understood as “attaching” or “associating” shapes to raw symbols (as, for example,
one “associates” sets to sort symbols in first-order logic) but rather as “moulding” mathematical objects into the appropriate shape. The mental picture evoked by this description should be that the pre-existing mathematical objects (which we want to show provide a model of the syntax) are being stretched into the shape of the syntax – and this is to be contrasted with the mental picture of being made to “fit” into the syntax, e.g. as a relation symbol is made to “fit” into a subset that interprets it.

A surrogate for the set-up I have just described already exists in set theory, in the usual way of carrying out the semantics for type theories. The raw syntax of type theories is molded into the shape of a contextual category and then homomorphisms from that contextual category into other (appropriately structured) contextual categories are regarded as models of the type theory. The contextual category acts as the intermediary, a kind of entity that is neither syntax nor semantics, but which contains the essential elements of both.\(^{53}\) It is possible, at this point, to regard such a contextual category as the syntax itself, i.e. identify the type theory with the contextual category it specifies.\(^{54}\) If one wants to avoid talk of mental pictures and shapes altogether and one is willing not to worry about the fact that contextual categories are first-order theories, then one may simply take my argument to be that the syntax of HoTTs should be identified with the contextual categories they specify.

This perspective suffices to answer the weak form of the PHA but not, to my mind, the strong one. And to answer the PHA in its strongest form, one must adopt a perspective on contextual categories that does not regard them as first-order theories. The way that I have argued this should be done is by regarding the syntax of HoTT as itself comprised

---

\(^{53}\)Lawvere and Makkai refer to the these kinds of entities, syntactic categories, as lying at the “conceptual level” between syntax and semantics. I am pointing out the exact same thing except I would rather call it the “spatial level”.

\(^{54}\)The philosophical implications of this idea, considered in more generality, is explored in [58], based on the program outlined in [57].
of shapes – and this point of view, I submit, is unique to HoTT, which is what allows it to answer the strong form of the PHA even if non-HoTT dependent type theories cannot.

We can thus condense this entire discussion into the following slogan: in UF, a piece of syntax \( S \) is satisfied by a model \( M \) if the shape of \( S \) is observable in \( M \). Although if both \( S \) and \( M \) are made precise in pre-existing frameworks then the above slogan can be seen merely as an unorthodox way of re-phrasing a pre-existing notion, what matters is that it has some intuitive notion of its own, and one which does not coincide with the “truth-making” intuition of first-order logic. Indeed, the notion of “truth” does not appear in the above slogan at all. And this, ultimately, is what allows UF to have a notion of satisfaction of its very own, independent from the “truth-assignment”-based notion of set theory. UF thus attains a level of autonomy unmatched by all previous categorical foundations of mathematics, thus allowing it to answer the PHA even in the very strong form that I have here considered. Whether this new notion of satisfaction enjoys the same kind of clarity, fundamentality or coherence as the traditional Tarskian notion is a philosophical question worthy of further exploration. I will say no more about it here.

Appendix: The Rules of 1UHoTT

In this appendix, I will give the full list of rules of the system of 1UHoTT that I have used as a “test subject” for the meaning explanation in Section 3.3. As in the main text I present them with “Shape” instead of “Type”. My presentation borrows elements from both [77] and [171].

Structural Rules.

\[
\begin{align*}
\Gamma \vdash A \text{ Shape} \\
\Gamma \vdash A = A \text{ Shape} \\
\end{align*}
\]

\[\text{Sym-shape-refl}\]
\[ \Gamma \vdash A = B \text{ Shape} \quad \Gamma \vdash B = A \text{ Shape} \]

\[ \text{Sym-shape-sym} \]

\[ \Gamma \vdash A = B \text{ Shape} \quad \Gamma \vdash B = C \text{ Shape} \]

\[ \text{Sym-shape-tran} \]

\[ \Gamma \vdash A = C \text{ Shape} \]

\[ \Gamma \vdash a: A \]

\[ \text{Sym-point-refl} \]

\[ \Gamma \vdash a = a: A \]

\[ \Gamma \vdash a = b: A \]

\[ \text{Sym-point-sym} \]

\[ \Gamma \vdash b = a: A \]

\[ \Gamma \vdash a = b: A \quad \Gamma \vdash b = c: A \]

\[ \text{Sym-point-tran} \]

\[ \Gamma \vdash a = c: A \]

\[ \Gamma \vdash a: A \quad \Gamma \vdash A = B \text{ Shape} \]

\[ \text{Sym-tran-1} \]

\[ \Gamma \vdash a: B \]

\[ \Gamma \vdash a = b: A \quad \Gamma \vdash A = B \text{ Shape} \]

\[ \text{Sym-tran-2} \]

\[ \Gamma \vdash a = b: B \]
\[ \Gamma \vdash a : A \quad \Gamma, x : A, \Delta \vdash b : B \]
\[ \Gamma, \Delta[a/x] \vdash b[a/x] : B[a/x] \]  
\text{Subst}_1

\[ \Gamma \vdash a : A \quad \Gamma, x : A, \Delta \vdash b = c : B \]
\[ \Gamma, \Delta[a/x] \vdash b[a/x] = c[a/x] : B[a/x] \]  
\text{Subst}_2

\[ \Gamma \vdash A \text{ Shape} \quad \Gamma, \Delta \vdash b : B \]
\[ \Gamma, x : A, \Delta \vdash b : B \]  
\text{Wkg}_1

\[ \Gamma \vdash A \text{ Shape} \quad \Gamma, \Delta \vdash b = c : B \]
\[ \Gamma, x : A, \Delta \vdash b = c : B \]  
\text{Wkg}_2

\vdash A_1 \text{ Shape}

\[ x_1 : A_1 \vdash A_2 \text{ Shape} \]
\[ \vdots \]

\[ x_1 : A_1, \ldots, x_{n-1} : A_{n-1} \vdash A_n \text{ Shape} \]

\[ x_1 : A_1, \ldots, x_n : A_n \vdash x_j : A_j \]

\text{Vble}, 1 \leq j \leq n, i \neq j \Rightarrow x_i \neq x_j

\text{Rules for type constructors.} We omit the rules expressing that logical constructors
are preserved under symmetry (i.e. under judgmental equality).

\text{Σ-types.}

\[ \Gamma \vdash A \text{ Shape} \quad \Gamma, x : A \vdash B(x) \text{ Shape} \]
\[ \Gamma \vdash \Sigma_x B(x) \text{ Shape} \]  
\text{Σ-Form}
Γ ⊢ A Shape  Γ, x: A ⊢ B(x) Shape
\[\text{Σ-Intro}\]
\[\Gamma, x: A, y: B(x) \vdash \langle x, y \rangle : \sum_{x:A} B(x)\]

Γ, z: \sum_{x:A} B(x) ⊢ C Shape
\[\text{Σ-Elim}\]
\[\Gamma, x: A, y: B(x) \vdash d(x, y): C(\langle x, y \rangle)\]
\[\Gamma, z: \sum_{x:A} B(x) \vdash \text{split}_d(z): C(z)\]

Γ, z: \sum_{x:A} B(x) ⊢ C Shape
\[\text{Σ-Comp}\]
\[\Gamma, x: A, y: B(x) \vdash d(x, y): C(\langle x, y \rangle)\]
\[\Gamma, x: A, y: B(x) \vdash \text{split}_d(\langle x, y \rangle) = d(x, y): C(\langle x, y \rangle)\]

Π-types.

Γ ⊢ A Shape  Γ, x: A ⊢ B(x) Shape
\[\text{Π-Form}\]
\[\Gamma \vdash \Pi_{x:A} B(x) \text{ Shape}\]

Γ, x: A ⊢ b: B(x)  Γ ⊢ \lambda(x: A).b: \Pi_{x:A} B(x)
\[\text{Π-Intro}\]

Γ ⊢ \Pi_{x:A} B(x)\]
\[\Gamma \vdash a: A\]
\[\Gamma \vdash f(a): B[a/x]\]
\[\text{Π-Elim}\]

Γ, x: A ⊢ b: B(x)  Γ ⊢ a: A
\[\text{Π-Comp}\]
\[\Gamma \vdash (\lambda(x: A).b)(a) = b[a/x]: B[a/x]\]
Id-types.

\[\Gamma \vdash \text{A Shape} \quad \Gamma \vdash a : \text{A} \quad \Gamma \vdash b : \text{A}\]

\[\Gamma \vdash \text{Id}_A(a, b) \text{ Shape}\]

\[\Gamma \vdash \text{A Shape} \quad \Gamma \vdash a : \text{A}\]

\[\Gamma \vdash \text{refl}_a : \text{Id}_A(a, a)\]

\[\Gamma, x : \text{A}, y : \text{A}, p : \text{Id}_A(x, y) \vdash \text{C Shape}\]

\[\Gamma, z : \text{A} \vdash d(z) : \text{C}(z, z, \text{refl}_z)\]

\[\Gamma, x : \text{A}, y : \text{A}, p : \text{Id}_A(x, y) \vdash J_z.d(x, y, p) : \text{C}(x, y, p)\]

\[\Gamma, x : \text{A}, y : \text{A}, p : \text{Id}_A(x, y) \vdash \text{C Shape}\]

\[\Gamma, z : \text{A} \vdash d(z) : \text{C}(z, z, \text{refl}_z)\]

\[\Gamma, x : \text{A} \vdash J_z.d(x, x, \text{refl}_x) = d(x) : \text{C}(x, x, \text{refl}_x)\]

+-types.

\[\Gamma \vdash \text{A Shape} \quad \Gamma \vdash \text{B Shape}\]

\[\Gamma \vdash \text{A + B Shape}\]

\[\Gamma \vdash \text{A Shape} \quad \Gamma \vdash \text{B Shape}\]

\[\Gamma, x : \text{A} \vdash 1(x) : \text{A + B}\]

\[\Gamma \vdash \text{A Shape} \quad \Gamma \vdash \text{B Shape}\]

\[\Gamma, x : \text{A} \vdash r(x) : \text{A + B}\]
\[
\Gamma, z : A + B \vdash C(z) \text{ Shape}
\]
\[
\begin{array}{ll}
\Gamma, x : A \vdash d_1(x) : C(1(x)) & \Gamma, y : B \vdash d_2(y) : C(r(y)) \\
\hline
\Gamma, z : A + B \vdash c_{d_1,d_2}(z) : C(z)
\end{array}
\]

\[-\text{Elim}\]

\[
\begin{array}{ll}
\Gamma, z : A + B \vdash C(z) \text{ Shape}
\end{array}
\]
\[
\begin{array}{ll}
\Gamma, x : A \vdash d_1(x) : C(1(x)) & \Gamma, y : B \vdash d_2(y) : C(r(y)) \\
\hline
\Gamma, x : A \vdash c_{d_1,d_2}(1(x)) = d_1(x) : C(1(x))
\end{array}
\]

\[-\text{Comp-1}\]

\[
\begin{array}{ll}
\Gamma, z : A + B \vdash C(z) \text{ Shape}
\end{array}
\]
\[
\begin{array}{ll}
\Gamma, x : A \vdash d_1(x) : C(1(x)) & \Gamma, y : B \vdash d_2(y) : C(r(y)) \\
\hline
\Gamma, y : B \vdash c_{d_1,d_2}(r(y)) = d_2(y) : C(r(y))
\end{array}
\]

\[-\text{Comp-2}\]

0-type.

\[
\begin{array}{l}
\hline
\Gamma \vdash 0 \text{ Shape}
\end{array}
\]

0-Form

\[
\Gamma, x : 0 \vdash C(x) \text{ Shape}
\]
\[
\begin{array}{l}
\hline
\Gamma, x : 0 \vdash \text{efq} : C(x)
\end{array}
\]

0-Elim

198
1-type.

\[
\begin{align*}
&\Gamma \vdash 1 \text{ Shape} \\
\hline
\Gamma \vdash * : 1 \\
\hline
\Gamma, x : 1 \vdash \text{rec}_d(x) : C(x) \\
\hline
\Gamma \vdash \text{rec}_d(*) = d : C(*)
\end{align*}
\]

N-type.

\[
\begin{align*}
&\Gamma \vdash N \text{ Shape} \\
\hline
\hline
\Gamma, \vdash 0 : N \\
\hline
\Gamma \vdash n : N \\
\hline
\Gamma \vdash s(n) : N
\end{align*}
\]
\[
\Gamma, x : \mathbb{N} \vdash \text{C Shape} \quad \Gamma \vdash c_0 : C[0/x] \quad \Gamma, x : \mathbb{N}, y : C \vdash c_s : C[s(x)/x] \quad \Gamma \vdash n : \mathbb{N} \quad \text{N-Elim}
\]

\[
\Gamma \vdash \text{ind}_{c_0, c_s, n} : C[n/x]
\]

\[
\Gamma, x : \mathbb{N} \vdash \text{C Shape} \quad \Gamma \vdash c_0 : C[0/x] \quad \Gamma, x : \mathbb{N}, y : C \vdash c_s : C[s(x)/x] \quad \text{N-Comp-1}
\]

\[
\Gamma \vdash \text{ind}_{c_0, c_s, 0} = c_0 : C[0/x]
\]

\[
\Gamma, x : \mathbb{N} \vdash \text{C Shape} \quad \Gamma \vdash c_0 : C[0/x] \quad \Gamma, x : \mathbb{N}, y : C \vdash c_s : C[s(x)/x] \quad \Gamma \vdash n : \mathbb{N} \quad \text{N-Comp-2}
\]

\[
\Gamma \vdash \text{ind}_{c_0, c_s, s(n)} = c_s : C[s(n)/x]
\]

\text{Rules for the Universe.}

\[
\begin{array}{ccc}
\vdash \cal{U} \vdash \text{Shape} & \quad & x : \cal{U} \vdash \text{El}(x) \vdash \text{Shape} \\
\end{array}
\]

\[
\begin{array}{c}
\Gamma \vdash a : \cal{U} \quad \Gamma \vdash b : \cal{U} \\
\vdash \text{Univalence}
\end{array}
\]

\[
\Gamma \vdash u_{a,b} : \text{isequiv}(\text{idtoequiv}_{a,b})
\]

\[
\begin{array}{ccc}
\Gamma \vdash a : \cal{U} \quad \Gamma, x : \text{El}(a) \vdash b(x) : \cal{U} \\
\vdash \pi(a, x, b(x)) : \cal{U} \quad \text{Uni-\Pi-1}
\end{array}
\]
\[
\Gamma \vdash a : \mathcal{U} \quad \Gamma, x : \text{El}(a) \vdash b(x) : \mathcal{U} \\
\Gamma \vdash \text{El}(\pi(a, x.b(x))) = \prod_{x : \text{El}(a)} \text{El}(b(x)) \quad \text{Shape}
\]

\[
\Gamma \vdash a : \mathcal{U} \quad \Gamma, x : \text{El}(a) \vdash b(x) : \mathcal{U} \\
\Gamma \vdash \sigma(a, x.b(x)) : \mathcal{U} \\
\Gamma \vdash \text{El}(\sigma(a, x.b(x))) = \sum_{x : \text{El}(a)} \text{El}(b(x)) \quad \text{Shape}
\]

\[
\Gamma \vdash a : \mathcal{U} \quad \Gamma \vdash b : \text{El}(a) \quad \Gamma \vdash c : \text{El}(a) \\
\Gamma \vdash \text{id}_a(b, c) : \mathcal{U} \\
\Gamma \vdash \text{El}(\text{id}_a(b, c)) = \text{Id}_{\text{El}(a)}(b, c) \quad \text{Shape}
\]

\[
\vdash z : \mathcal{U} \\
\vdash \text{El}(z) = 0 \quad \text{Shape}
\]

\[
\vdash s : \mathcal{U} \\
\vdash \text{El}(s) = 1 \quad \text{Shape}
\]
\[
\Gamma \vdash a : \mathcal{U} \quad \Gamma \vdash b : \mathcal{U} \\
\hline
\Gamma \vdash a \oplus b : \mathcal{U} \\
\hline
\text{Uni-} -1
\]

\[
\Gamma \vdash a : \mathcal{U} \quad \Gamma \vdash b : \mathcal{U} \\
\hline
\Gamma \vdash \text{El}(a \oplus b) = \text{El}(a) + \text{El}(b) \quad \text{Shape}
\]

\[
\text{Uni-} -2
\]

\[
\Gamma \vdash N : \mathcal{U} \quad \Gamma \vdash \text{El}(N) = N \quad \text{Shape}
\]

\[
\text{Uni-} \text{N-1} \quad \text{Uni-} \text{N-2}
\]
CHAPTER 4

On Alternative Structuralist Foundations

The Univalent Foundations certainly are not the only candidate structuralist foundations and in this Chapter I compare and contrast them to alternative proposals. In Section 4.1 I review several categorical foundations that have been suggested and explain in what sense Makkai’s Type-Theoretic Categorical Foundations of Mathematics (TTCFM) emerges as the most likely contender to UF. In Section 4.2 I will compare UF and TTCFM on several issues, and conclude that the most important point of comparison concerns their intended semantics (weak $\infty$-categories vs. weak $\infty$-groupoids). The rest of the Chapter is then devoted to producing an elementary, non-technical argument to choose between these two intended semantics. In Section 4.4 I will show how the dispute reduces to the question of “What is a higher-level set?” by introducing the relevant ideas from category theory and examine possible methods to pick between two different answers, groupoids or categories. Finally, in Section 4.5 I will offer an argument in favour of groupoids by distinguishing between organizational and spatial aspects of groupoids and categories.\(^1\)

4.1. Categorical Foundations and Makkai’s TTCFM

The story of categorical foundations of mathematics begins with a very basic observation in Lawvere’s work on ETCS [95]: category theory could be used to frame set theory, rather than category theory being one more theory defined inside set theory. This general guiding idea has given rise, over the past half-century, to a plethora of suggestions as to

\(^1\)An early version of this Chapter was presented in a Symposium on the Foundations of Mathematics (“SoTFOM”) in Birkbeck College, London, January 2015.
what would constitute a *categorical foundation of mathematics*. Not all of them have been incompatible with set-theoretic thinking; but almost all of them have argued for a revision, of varying magnitude, to the standard picture of the foundations of mathematics encapsulated by set theory and ZFC. To decide whether any of these categorical foundations can compete with UF, I will now examine various proposals for categorical foundations that have emerged since Lawvere.\(^2\) Throughout, I will employ the minimalist picture of foundations sketched in Section 2.1 and my use of the terms “basic language”, “theory” and “universe of objects” as they pertain to a particular foundational proposal will coincide with the one described there.

As just suggested, historically the first approach to categorical foundations was that of Lawvere, outlined in his pioneering paper [95], where he gave a list of first-order axioms meant to describe the category \textbf{Set} of sets; his perspective on foundations is further sharpened in subsequent work, especially in [93]. In this foundational proposal, the basic language is first-order logic, but instead of a primitive membership relation “\(\in\)” we have a primitive notion of a “function” with domains and codomains. As was discussed in Chapter 2, this approach allows us to obtain a more “structural” (as opposed to “extensional”) approach to sets, where priority is given to how these sets relate one to the other, rather than to whether or not they contain this or that element.

The defining feature of this approach is that it maintains first-order logic as its basic language. Variations of ETCS (cf. [96, 143] for modern presentations) as well as Lawvere’s proposal for a *Category of Categories as a Foundation* (CCAF) [94, 130] emerge as the most prominent proposals in this strain. Such proposals have been savaged by various philosophers (cf. [61, 162]) and logicians (cf. [41]) and I believe the main thrust of this criticism to be correct: “elementary” categorical foundations do become parasitic

\(^2\)My breakdown of the history of categorical foundations is very much influenced by the work of Marquis and Landry in [91] and the reader is referred there for more detail.
on set theory on some level, whether ontological, mathematical or logical. This leads to
the following dilemma: either one simply calls theories like ETCS “set theories”\(^3\) in which
case category theory really plays no fundamental role, or one changes one’s conception of
the foundations of mathematics towards an account where “foundation” is understood as
a kind of framework in which mathematical practice can be fruitfully organized. Given
the minimalist conception of the foundations of mathematics that I am working with, the
second horn is not an option; and given what was said about the capacity of theories like
ETCS to improve upon (SFOM) as outlined in Chapter 2, neither is the first. I do not
therefore consider any structuralist foundations that emerge from this strain as worthy
competitors to UF.

Moving on, there are some proposals for categorical foundations that do not fit the
minimalist picture, which by my lights is reason enough to disregard them as serious
contenders to UF. I will briefly describe them, explaining in each case how they fail to
fit the minimalist picture. Mac Lane \([100–102, 104]\) propounded a sort of foundation-
ally nihilistic view comprised of a hodgepodge of various philosophical and mathematical
attitudes: form and function should take precedence over substance and essence; clarity
and usefulness should be emphasized over ontology; the idea of structure figures promi-
nently in order to capture the “many-splendored” nature of the objects of mathematics;
any kind of foundationalism is rejected since mathematics is a “protean science”. Mac
Lane’s vague, colourful and often inspiring suggestions certainly amount to a particular
attitude one can take towards mathematics, even a certain kind of aesthetic constraining
the methods one may use, but they in no way pin down a specific formal system meant
to describe some fixed universe of objects. Mac Lane’s view is thus best described as a

\(^3\)Strictly speaking, this is exactly what they are, since the variables in their first-order axiomatizations
are meant to be interpreted as sets.
framework, not as a foundation, and category theory plays almost an incidental role in it.

Similarly, John Bell [16–18] has advocated in favour of topos theory as a kind of “global” framework in which many different “local” ways of doing mathematics all live together, connected by the logical and geometric morphisms that connect toposes. Though Bell was initially not very hostile to the idea of a set-theoretic foundation to underlie such a framework (e.g. in [16]) he does move away from this perspective eventually, and becomes more of a nihilist along the lines of Mac Lane; in particular, he rejects the idea that a foundation should consist of one single theory over a single basic language. At its best, therefore, Bell’s framework is a kind of topos-theoretic justification for logical pluralism (based on the fact that different toposes encode different logics) and not a foundation in the sense of the minimalist picture.

On the other hand, Lambek [34, 86, 88] and collaborators define a foundational proposal in which the universe of objects of mathematics is given by the “free syntactic topos”. On the surface, this looks like a proposal that would fit the minimalist picture. Its basic language is (intuitionistic) simple type theory, its theory is comprised of whatever extra structure is needed for such a type theory to generate a topos, and its universe of objects is taken to be the topos generated by this syntax. The problem with Lambek’s picture is that actual mathematics is not encoded in terms of the objects of a “free syntactic topos”. Rather, these deductions are put into the very definition of the syntactic free topos (it is what makes it “syntactic” to begin with). As such, the universe of objects in Lambek’s proposal is defined only in terms of the very process of encoding that any foundation, on the minimalist picture, ought to enable. It cannot therefore be regarded as a foundational proposal in my sense, regardless even of its categorical character. But this approach does
define an interesting view in the philosophy of mathematics, one that Lambek and Scott coin “constructive nominalism”, and which perhaps merits further exploration.

What remains is the proposal of Makkai, originally called Structuralist Foundations of Abstract Mathematics (SFAM) [110,111] and in its latest instantiation called Type-Theoretic Categorical Foundations of Mathematics (TTCFM) [113]. According to the minimalist picture (into which, unlike other proposals, it straightforwardly fits) TTCFM is a foundational proposal consisting of the following elements:

\[
\text{FOLDS} \rightarrow \text{MultCat} \rightarrow \infty\text{Cat}
\]

I will now briefly describe each of these components, and assess their state of development. First Order Logic with Dependent Sorts (FOLDS) was invented by Makkai with the specific intention of providing the syntactic underpinning of a structuralist foundation of mathematics. In broad brushstrokes, Makkai’s intention was to come up with a syntax such that every grammatically well-formed sentence is invariant under a (semantically-defined) notion of equivalence that happens to coincide with criteria of identity defined in mathematical practice (e.g. isomorphism of structured sets).\(^4\) As the name suggests, FOLDS is to be thought of as a “first-order” dependent type theory very much in the style of MLTT, but it can also be viewed as a fragment of first-order logic since there is a canonical translation of FOLDS signatures into first-order signatures and FOLDS formulas into first-order formulas.\(^5\) It also closely resembles many “first-order type theories” that have been studied in different contexts and for different purposes, e.g. the bracket type theories of [10], the minimal type theories of [108,109], the predicative type theories of [139] and the logic-enriched type theories of [49]. Very much related is

\(^4\)A detailed exposition of FOLDS can be found in [110] and more recently in [113], and an account of its philosophical motivation can be found in [111] and [122]. I will also have much more to say about FOLDS in Chapter 5.

\(^5\)This is no longer true for \(n\)-logic as I define it in Chapter 5.
also the work of Cartmell [28] which marks the beginning of the idea of giving categorical semantics to MLTT.

There are three main results that make FOLDS a suitable basic language of a structuralist foundation. Firstly, that FOLDS is at least expressive enough to be able to axiomatize the structures we are interested in and also to express the properties of those structures that are of interest in practice. Secondly, that FOLDS-equivalence (the correct notion of “elementary equivalence” for FOLDS-structures) corresponds to the usual notions of equivalence when specialized to familiar cases. For example, that it specializes to categorical equivalence in the case of categories and, as we shall see in Section 4.2, to homotopy equivalence for an appropriate choice of signature. Thirdly, that for any first-order sentence $\phi$ that is invariant under FOLDS-equivalence there is a FOLDS sentence $\tilde{\phi}$ equivalent (in an appropriate sense) to $\phi$. (If the first fact shows that FOLDS is minimally adequate, this fact shows that it is also maximally so.) To quote Makkai:

> [F]or any general first-order statement $\phi$ that is formulated in some language possibly extending $\mathcal{L}$, if $\phi$ is (universally) invariant under FOLDS $\mathcal{L}$-equivalence, then there is a statement $\psi$ written down in FOLDS using the vocabulary $\mathcal{L}$ which is equivalent, for all structures under consideration, to the original $\phi$. ([112], p. 14)

All three of these facts have been convincingly established by Makkai in several of his (still unpublished) writings on FOLDS (most of them in [110]).

Interesting though it may be, FOLDS is not a foundational proposal in itself (any more than first-order logic without the axioms of set theory is a foundational proposal). It is merely the basic language of one. So what is the theory (i.e. component (2)) expressed

---

6In the case of categories, for instance, he has shown that Set-based categories are axiomatizable as FOLDS structures and that FOLDS equivalence corresponds to the usual notion of categorical equivalence in that setting. The interested reader may find precise statements and proofs of these results in [111] and [110].
in this basic language that axiomatizes the ∞-category of ∞-categories? Unfortunately, such a theory does not yet exist in a usable form. What does exist is the so-called multitopic definition of ∞-categories provided by Makkai, Street and Powers in the series of papers [114–116]. This work remains the only fully worked-out definition of ∞-categories in the form in which TTCFM would require them to exist. But although there are elements of FOLDS in the multitopic definition, it is still not a proper formal definition (in the sense of a list of axioms) in FOLDS. Insofar as no such axiomatization has been given, TTCFM remains an ambitious proposal, but not a completed one.

However, TTCFM’s intended semantics, namely ∞-categories, make perfect sense in mathematical practice and indeed are of interest to practitioners. For example, these structures have found increasing relevance in algebraic geometry [99] and even in number theory as a possible way forward in the so-called “geometric Langlands” program. As their respective names suggest, there is a close relationship between ∞-groupoids and ∞-categories. Interestingly, there seems to be no immediate reason to pick ∞-groupoids over ∞-categories with respect to the question of whether they can play the role of the basic objects of a structuralist foundation for mathematics. Thus, if we put to the side the issue of a formal axiomatization (which TTCFM lacks), TTCFM raises a very interesting challenge to UF: why not take ∞Cat (rather than ∞Gpd) as the basic objects of our structuralist foundations? The greater part of this Chapter will be concerned with resolving exactly this question. But first, we compare UF and TTCFM on two further issues.

### 4.2. Comparing UF and TTCFM

All in all, even despite the lack of a formal theory, TTCFM emerges as the most interesting competitor to UF. In this Section I will examine key differences between these

---

7For a more leisurely introduction to the ideas contained in these papers see Makkai [112].
two proposals by comparing them on two fronts: their approach to invariance and their capacity to carry out their own metatheory.

4.2.1. Language and Identity. UF and TTCFM provide examples of two distinct ways one may set about defining a structuralist foundation of mathematics:

(a) By placing restrictions on the expressive power of a given language.

(b) By postulating a “transport principle” that allows one to transfer properties along equivalences.

The first approach (a) involves fixing a syntax $\mathcal{S}$ together with an appropriate formal system first, demonstrating that the set-up is sufficiently expressive to serve as a foundation for mathematics, and then defining a notion of $\mathcal{S}$-equivalence such that nothing can be expressed in $\mathcal{S}$ that is true of one “structure” (the term is used informally here) but not of another structure $\mathcal{S}$-equivalent to it. This approach fixes a language first and then a notion of equivalence. We may call it the language-first approach. This is the approach taken by TTCFM, where $\mathcal{S}$ is FOLDS and $\mathcal{S}$-equivalence is FOLDS-equivalence.

The second approach (b) involves first settling on a notion of equivalence and then building a syntax $\mathcal{S}$ that does justice to this already-chosen notion of equivalence in the following sense: firstly, that $\mathcal{S}$ formalizes the pre-formal notion of equivalence by some formal relation $\equiv_{\mathcal{S}}$; secondly, and more importantly, that nothing is expressible in $\mathcal{S}$ that is true of one “structure” but not of one that is equivalent $\mathcal{S}$ to it. We may call this the identity-first approach. In the case of UF, $\mathcal{S}$ is a dependent type theory, equivalence $\equiv_{\mathcal{S}}$ is $h$-equivalence and the axiom of univalence ensures that nothing can be true of a structure without also being true of a structure $h$-equivalent to it by ensuring that any property of one structure can be transferred to any other $h$-equivalent structure.

The philosophically relevant data point here is that these two approaches seem to be associated to distinct views on how exactly we create mathematical meaning. On the
one hand, the language-first approach presupposes that mathematical meaning is determined in all its essentials by mathematical practice – and that the task of structuralist foundations is to capture this meaning in a uniform language that mirrors practice as much as possible. The relevant notion of equivalence between mathematical structures is then determined \textit{a posteriori} as the notion which respects exactly the distinctions made by this language. The closest analogue of such an approach in traditional foundations would then resemble something like a pragmatic, Carnapian view on the meaning of mathematical statements as determined by the acceptance of a linguistic framework in which mathematical questions are posed (as opposed to being determined by how such a linguistic framework corresponds to “reality”, whatever that may be). And just as Carnap insisted that the acceptance or rejection of a particular linguistic framework is a \textit{pragmatic} question, so should the acceptance or rejection of a language-first structuralist foundation come down to \textit{pragmatic} considerations, viz. how well the given language corresponds to the language of mathematical practice.

On the other hand, the identity-first approach pre-supposes that mathematical meaning is determined by first selecting a universe of objects (in the case of UF, homotopy types) together with a criterion of identity for them (in the case of UF, homotopy equivalence) and only then determining the particulars of the formal system (i.e. the language) that is going to formalize it. In the identity-first approach, therefore, it is the invariant properties that are determined \textit{a posteriori} as those properties that are invariant under an \textit{a priori} determined criterion of identity. Perhaps the closest historical analogue to such an approach in traditional foundations is the Fregean approach to defining numbers in terms of a previously-given notion of numerical identity.\textsuperscript{8}

\textsuperscript{8}The most relevant passage is perhaps from §63 of the \textit{Grundlagen}:

\ldots[F]or us, the concept of number has not yet been defined, but rather is to be determined by means of our definition of numerical identity. We intend to reconstruct the content of judgments interpretable as expressing identities each side of which is a number. We do not, therefore, want to define equality especially for this instance,
All this seems to be pointing towards some kind of realist/anti-realist dichotomy with respect to mathematical objects with the language-first approach corresponding to anti-realism and the identity-first approach corresponding to realism. In more detail, the language-first approach sounds closer to a verificationist (e.g. along the lines of Dummett) view in the sense that it seems to be saying that mathematical meaning is determined by the use mathematicians make of mathematical statements. The identity-first approach sounds closer to some kind of Platonism (e.g. about $\infty$-groupoids) in the sense that what it is for something to be a grammatical mathematical property is determined independently of mathematical practice in some abstract realm (e.g. the realm of homotopy types). Does this mean that TTCFM (or any language-first approach to (SFOM)) and UF (or any identity-first approach to (SFOM)) commits one to anti-realism or realism respectively in the philosophy of mathematics?

I do not think so. Even though there is certainly some philosophical connection (or affinity) between TTCFM and anti-realist (or Carnapian) views on the one hand and UF and realist (or Fregean) views on the other, we need to remember that the focus here is on the foundations of mathematics, and not necessarily on their ontology or epistemology. Particular choices of foundations for mathematics in no way commit us to particular philosophical views on their ontology or epistemology. There are affinities, to be sure, between certain choices of foundations and certain philosophical views (e.g. between intuitionistic ZF and intuitionism) but such affinities hardly add up to full-blown philosophical commitment. I do not think this is particularly controversial: I am merely expressing the fact that, for example, developing a foundational set theory in the manner

but we wish rather, by means of the already familiar concept of equality, to determine that which is to be considered equal.

Replacing “equality” for “homotopy equivalence” in the last sentence above would give exactly the spirit in which the basic objects (homotopy types) of UF are determined.
of ZFC or NBG or NF does not thereby automatically commit you to a certain view in the philosophy of mathematics.\(^9\)

In the case of structuralist foundations, this separation between choice of foundation and commitment to a particular philosophical view is even more pronounced than in the case of traditional foundations.\(^10\) In the case of traditional foundations some kind of “ultimate” philosophical justification for particular formal features (e.g. for large cardinal axioms something along the lines of \([105, 106]\)) may indeed prove to be an important feature, in particular with respect to justifying the role of traditional foundations as “benchmarks of consistency.” But in the case of structuralist foundations, the justificatory spade is turned at the level of mathematical practice. And mathematical practice is in no way constrained by the philosophy of mathematics.\(^11\) Therefore, any view in the philosophy of mathematics should at least be (dialectically) compatible with any choice of structuralist foundation (identity-first, language-first or otherwise).\(^12\)

However, there is a curious and interesting fact that emerges from this linguistic comparison of UF and TTCFM, namely that their criteria of identity, though arrived at differently, actually coincide. This is made precise by the following theorem, based on an observation of Makkai in \([112]\). The theorem is made precise and proved in an Appendix to this chapter – I will here state its “informal” version, which captures its intuitive content:

**Theorem 4.2.1.** Two \(\infty\)-groupoids are homotopy equivalent if and only if they are **FOLDS-equivalent** when we consider them without identities.

---

\(^9\)It *does* of course commit you to regarding certain features of a foundation as more essential than others, e.g. having a distinct sort symbol for proper classes or not or having a set of all sets.

\(^10\)I am using the terms “traditional” and “structuralist” in the sense expounded in Chapter 3.

\(^11\)Nor, it has to be said, is the philosophy of mathematics usually constrained by the practice of mathematics. Corfield [32] is a dissenting voice in that regard.

\(^12\)I am not sure such a strong statement also holds in the case of traditional foundations (which I am not concerned with here). For example, it seems to me that choosing NF as one’s preferred (“true”) foundation seems to me to restrict one dialectically to a certain range of philosophical views on mathematics.
Theorem 4.2.1 says that homotopy equivalence can be recovered as a special case of FOLDS-equivalence. Thus, one way of interpreting this result is as follows: homotopy equivalence emerges naturally as a notion with which to compare mathematical objects if one cares about systematically eliminating statements that are of no use in mathematical practice. In other words, there is something irreducibly *logical* about homotopy equivalence, since the notion emerges naturally from considerations that have nothing to do with topology. It thus demonstrates that the notion of homotopy equivalence, on which UF relies, can be arrived at by *purely logical* considerations, such as the ones Makkai was concerned with when he was trying to find a language that made no distinctions that mathematicians themselves wouldn’t make. Therefore, whether language-first or identity-first, the notion of homotopy equivalence seems to naturally emerge as a useful notion in structuralist foundations.

4.2.2. Metatheory. Any foundation worth its name should be able to provide a framework in which to carry out not only its own metatheory but also metamathematical investigations in general.\(^{13}\) Set theory has been monumentally successful in this task: set theories like ZF(C) are perfectly capable of expressing what it is for something to be a model of ZF(C) and then proceed to carry out metatheoretical investigations on ZF(C) in this manner.\(^{14,15}\)

But the situation with UF is not at all as simple as that. There is no currently available way to say “Let \(\mathcal{U}\) be a model of UF ... ” *inside* UF. This is because, roughly, in order to express that a certain type is a model of UF one needs to express that it satisfies infinitely complicated coherence conditions.\(^{16}\) How to do the latter, indeed, is

\(^{13}\)Indeed, a version of this demand is exactly the PHA challenge considered in Chapter 3.

\(^{14}\)Of course the *existence* of such a model cannot be proven in ZF by Gödel’s second incompleteness theorem.

\(^{15}\)Gödel and Cohen’s celebrated proofs (which, when combined, yield the independence of the continuum hypothesis from the axioms of ZFC) take place in ZF, even though they concern formal theories at least as strong as ZF itself.

\(^{16}\)See \([163]\) for a nice and (somewhat) accessible technical overview of this problem.
one of the most active current areas of research in HoTT. Indeed it is still very much an open question whether it is at all possible that HoTT as formalized in [171] (or 1UHoTT as in Chapter 3) can serve as its own metatheory.

The reason for this, roughly, is that the native identity in HoTT (namely identity types) does not behave as we expect it to behave when applied to raw syntax. To illustrate, take the case of variables. As pieces of syntax, variables are necessarily understood as (decidable) sets of variables. (This also goes for other syntactic entities, but the problem is most vivid in the case of variables.) In other words, we need to differentiate between “x” and “y” (regarded as individual symbols) exactly like we differentiate between elements of sets, i.e. we care only about whether they are equal or not (as symbols). We seem to have no use for questions like: “In how many ways could two distinct symbols for variables be equal?” But in systems like HoTT the following conflict emerges: the variables in the system are meant to denote entities that do not have “set-like” identity conditions; namely they denote points in (possibly higher) groupoids. But if we are to do the metatheory of such a system inside such a system then we need to use the “native” identity to differentiate variables and it turns out that we need to differentiate variables far more finely than what this “native” notion, in general, would allow.

On the other hand, it is far clearer how one would carry out the metatheory of TTCFM inside of TTCFM than it is clear how one could carry out the metatheory of UF inside UF. This has little to do with the specifics of TTCFM and more with the fact that TTCFM is based on a proof-irrelevant syntax (FOLDS).\footnote{It is true of course that the semantics of FOLDS are (usually) developed in a set-theoretic framework. But the fact that completeness and soundness has been proven for these semantics (cf. [110]) using this set-based metatheory does not mean that FOLDS is deep-down about a universe of entities that can only be described by an axiomatic set theory. Indeed in [113], Makkai makes exactly this point: It is important to note that the meta-theory of TTCFM does depend on established set theory. However, its formalized version, being explicit and elementary, that is, first-order, stands up on its own, without anything “prior to it”. These is no} Thus, the theory of
∞-categories in TTCFM will be able to serve as its own metatheory essentially in exactly the same way that ZF serves as its own metatheory: encoding the syntax of FOLDS and stating what it is for (a certain variable) to be a model of certain FOLDS sentences in TTCFM\(^{18}\) will happen just as it happens in set theory. This is because of the proof-irrelevance of the syntax which allows us to separate it completely from the semantics. More precisely, because TTCFM is built on a proof-irrelevant syntax there is a strong separation between axioms and language, i.e. between “structures” (which the object language can talk about directly) and “propositions” (which the object language cannot talk about directly). But UF (and intensional MLTT in general) does not fit the same mold: one does not first define a language and then state axioms in that language. One simply defines the language, i.e. the deductive system and the “axioms” are the language itself. As discussed at length in Chapter 3, a “model” of UF is not some kind of structure (or process) that makes certain propositions true, but rather a some kind of structure that can “mimic” all the moves one is allowed to make in the given deductive system.

This brings us to what may prove a fundamental issue in any kind of approach to alternative foundations (structuralist or otherwise): it seems much easier (both technically and conceptually) for formal systems to serve as their own metatheories when these systems are defined over a proof-irrelevant syntax (e.g. FOL\(_=\) or FOLDS) than when they are defined over a proof-relevant syntax (e.g. (intensional) MLTT.)

4.3. (∞-)Groupoids Versus (∞-)Categories

I will now move to the most crucial debate between UF and TTCFM: their intended semantics. Is there any purely philosophical way to choose between \(\infty\text{Gpd}\) and \(\infty\text{Cat}\) contest intended with established set theory: I am proposing an alternative, not an exclusive, foundational scheme. The system of TTCFM [...] can also serve as the basis of significant metatheoretical considerations. ([113], p. 4)

\(^{18}\)Assuming, of course, that TTCFM is strong enough to contain a “natural number object” that models some strong enough theory of arithmetic.
with respect to their suitability to play the role of basic objects of a structuralist founda-
tion of mathematics? The issue is much more general than the specific debate between \( \infty\text{-Grpd} \) and \( \infty\text{Cat} \) might indicate. Essentially, the debate is about the question: what is a higher-level thing?

The notion of “higher” that is relevant here comes from category theory. It is the notion of categorification (or groupoidification/homotopification.) \( Y \) is a higher-level (or next-level) \( X \) if \( Y \) is a categorification of \( X \). If recent developments in theoretical physics\(^{19}\) and the foundations of mathematics are anything to go by this notion will become increasingly relevant to philosophers in the coming years. But even if one is left unmoved by developments in the foundations of mathematics and theoretical physics, it is worth recalling that many deep philosophical questions are about determining whether a certain phenomenon is of a “higher-level” than another: are mental states “higher-level” brain states? Is colour a “higher-level” property of particle arrangements? As such, a purely philosophical investigation of this new category-theoretic notion of “higher-level” seems to be in order. I will therefore undertake it in a much more general setting than the foundations of mathematics.

To do this, we must first ask: intuitively, what are \( \infty \)-categories and \( \infty \)-groupoids? Roughly, they are mathematical structures with “maximally weak” identity conditions between their elements: to ask whether any two of their elements are “equal” amounts to finding a “process” which establishes this equality; and verifying that process requires another process – and so on to infinity. One can illustrate the idea with a little fable: Suppose you have two city-states, say Athens and Sparta (our two “elements”). Suppose that Athens and Sparta want to forge an alliance to fend off the invading Persian army (“forging an alliance” being the relevant notion of “equality”). In order to do so, Athens

\(^{19}\)For example, work on higher gauge theories, e.g. [158].
picks one of its citizens to carry a message to Sparta, asking for an alliance (this is the “process” that establishes the “equality”). The messenger runs to Sparta, delivers the message and then runs back to Athens. He announces that the Spartans have accepted. The Athenians, not certain they can trust the messenger, send him back to Sparta but this time accompanied by someone they trust even more, perhaps a higher-ranking soldier. Predictably, both messengers come back and announce that the Spartans accept the Athenian offer. But after another short-lived wave of relief, suspicion once again fills the hearts of the philosophizing Athenians. Why should they trust the second soldier? The Spartans could very well have bribed him too. The solution is to send an even more trustworthy citizen to accompany the other two. And so on: the suspicion of the Athenians can never be fully allayed. At some point an arbitrary decision will have to be made: someone will be deemed sufficiently trustworthy (perhaps Pericles himself) and the alliance will be forged.

But if this arbitrary decision is never taken, then the process of verification will never end. Mathematically, we can very well imagine postponing this decision indefinitely (i.e. requiring an infinite number of messengers before we can be sure that an alliance has been forged). ∞-categories and ∞-groupoids are then exactly such mathematical structures, in which a decision about whether any two of their elements are “equal” can be indefinitely postponed.

Needless to say, to explicitly describe such structures involves non-trivial mathematics. Nevertheless, the main idea behind a weak ∞-category/groupoid is fairly easy to illustrate to anyone with a passing acquaintance of category theory. An ∞-category, for example, is a category in which every category axiom holds only up to isomorphism where “isomorphism” is determined up to isomorphism and so on all the way to infinity.\footnote{Take the (left) unit law as an illustration. This says that for any object $A$ in a category there is an arrow $1_A: A \to A$ such that for any $f: B \to A$ we have $f = 1_A \circ f$. In a weak ∞-category this same law}

218
The simplicity of the basic idea ought to instill some hope that there is a purely philosophical way to choose between $\infty$-groupoids and $\infty$-categories as the basic objects of a structuralist foundation. What’s more, the question, I submit, reduces to the following far more tractable one: “What is a higher-level set?” This is because both $\infty$-groupoids and $\infty$-categories can be understood as being stratified in hierarchies generated by a single principle. Namely, we have $n$-groupoids and $n$-categories for $0 < n \leq \infty$ and the principle that moves us from $n$ to $n+1$ is in each case the same, for every $n$, as we shall see. Therefore, if we can decide the question at level 0 (assuming, as we will, that 0-groupoids and 0-categories are the same notion, viz. sets) then we have decided the question in the general case. The rest of this Chapter is therefore devoted to answering the question “What is a higher-level set?” – I will begin by trying to get some much-needed clarity on the relevant notion of “higher level”.

4.4. The Notion of “Higher-Level”

Consider a bowl filled with raffle tickets. Someone is assigned the task of shuffling the tickets. For certain questions we might want to ask of the bowl, the extra information about who shuffled it and how will not be relevant. For example, if we want to ask: “How many tickets are in the bowl?” or “Who won the raffle?” For certain other questions, information about who did the shuffling will become relevant: “Was the shuffling

would now instead stipulate the existence of an isomorphism

$$f \cong 1_A \circ f$$

and an “isomorphism” here would be understood as a 2-arrow (an arrow between arrows)

$$\alpha: f \Rightarrow 1_A \circ f$$

such that there exists an “inverse” 2-arrow, i.e.

$$\alpha^{-1}: 1_A \circ f \Rightarrow f$$

such that, once again, $\alpha \circ \alpha^{-1} \cong 1_A$ and $\alpha^{-1} \circ \alpha \cong 1_f$ where these isomorphisms are again determined up to 3-arrows and so on to infinity. And exactly analogously, $\infty$-groupoids can be thought of as $\infty$-categories in which every higher arrow is invertible (up to isomorphism, up to higher isomorphism, etc.).

These are not, however, generative hierarchies in the same sense that the cumulative hierarchy is generated by a certain number of operations.
thorough?” or “Why is the winning ticket smudged with barbecue sauce?” If we are interested in asking such questions then it is better to think of our object of interest not as a bowl of raffle tickets but as a bowl of raffle tickets together with someone assigned to shuffle it. Then the relation of the bowl-and-shuffler to the bowl is an illustration of the relation I am interested in: the bowl-and-shuffler is a higher-level bowl.

Consider a laptop not connected to the Internet. Then, suddenly, its owner decides to connect the laptop to the Internet. The device she now has before her is not (just) her laptop. In fact, what she has before her is the whole Internet. In being connected to a network of other laptops (and servers) the owner has access to a device over and above the laptop before her. The relation of (offline) laptop to Internet is also an illustration of the relation I am interested in: the Internet is a higher-level laptop.

These illustrations tack on to familiar intuitions. In moving from bowl to bowl-and-shuffler what we are doing is moving from an object to an object-with-symmetries. Namely, the “higher-level”-ness of the bowl-and-shuffler consists in keeping track of certain symmetries of the object that we had initially deemed irrelevant. This is entirely analogous to moving from a spacetime $\mathcal{S}$ (e.g. $\mathbb{R}^4$) regarded in isolation to a spacetime-with-symmetry group pair $(\mathcal{S}, G)$ (e.g. $(\mathbb{R}^4, \mathcal{L})$ where $\mathcal{L}$ is the Lorentz group). On the other hand, in moving from the laptop to the Internet what we are doing is moving from an object to a structured network of similar objects. Namely, the “higher-level”-ness of the Internet consists in being a collection of connected laptops (and servers). There is no shortage of controversial examples of this relation: a mind is a collection of connected neurons; colour is the way certain molecules hang together. And so on.

Since I want to view both these illustrations as examples of a similar phenomenon, a grammatical clarification is in order. When I speak of a “higher-level $X$” I take $X$ to designate the class of things to which it belongs and not necessarily some specific
individual in that class. A “higher-level apricot” does not ask for the higher-level version of the particular apricot on my table, but of the species “apricot” of which the apricot on my table is a particular instance. Of course, it might happen that occasionally – as with the raffle-bowl – a higher-level \(X\) might make sense even for a specific \(X\). But it need not do, as the example with the laptops hopefully illustrates.\(^{22}\)

Both notions of “higher-level” that the above examples are meant to illustrate originate from category theory and only properly make sense in a mathematical setting. So let us put apricots and neurons to the side for now and offer some concrete mathematical examples:

(1) An abelian category is a \textit{higher-level} abelian group.

(2) A monoidal category is a \textit{higher-level} monoid.

Both examples are categorifications in their own unique way. An abelian category is a category that resembles the category \(\text{Ab}\) of abelian groups. A monoidal category is a category together with a multiplication operation (usually denoted “\(\otimes\)”) that behaves just like the multiplication in a monoid except its properties are defined “up to natural isomorphism.” Of these examples, it is only the second (monoids to monoidal categories) that is properly speaking a \textit{categorification}. But both examples exhibit the two essential features of categorification that I want to focus on below:

(1) Moving from a (structured) object (e.g. abelian group, monoid) to a (structured) \textit{collection} of objects (e.g. abelian category, monoidal category).

(2) Weakening identities (e.g. \(1 \times x = x, \ g \circ f = h\)) to natural isomorphisms (e.g. \(\alpha_x: 1 \otimes x \cong x, \ g \circ f \cong h\)).

\(^{22}\)For the purposes of this Chapter, it is safe to assume the question “What is a higher-level \(X\)?” is an abbreviated form of the much more cacophonous: “What is type of thing that can be regarded as the higher-level version of the type of thing that \(X\) is?”

221
The details of these mathematical examples are not at all essential to what I have
to say below. Although it would be interesting to try and come up once and for all
with a mathematical definition of categorification, this is not what I want to do here.\textsuperscript{23}
Rather, as indicated above, in this Chapter I will focus on the simplest of mathematical
examples, viz. that of a \textit{set}. What is the higher-level version of a set, when “higher-
level” is understood in the above-sketched manner? Immediately, this question raises
difficulties that other mathematical examples do not. Firstly, if we want to categorify
a set by moving to a collection of sets, we have no pre-existing “structure” to use as a
guide to determine what structure to impose on this collection as we did, for instance,
in the case of monoids where we had a multiplication operation. Secondly, what kind of
identities does a set satisfy that we could weaken to isomorphisms?

As it turns out, each of these concerns leads naturally to distinct answers to the
question of what is a higher-level set: categories or groupoids. Let me recall the formal
definitions of a category and a groupoid, although nothing I say hinges on a precise
understanding of these definitions. Indeed, I will also develop pre-formal definitions for
both notions below and the crux of my argument will depend only on those. Nevertheless,
for the sake of completeness, it is helpful to lay them out.

\textbf{Definition 4.4.1.} A \textbf{category} \( \mathcal{C} \) consists of a collection of objects \( \text{Ob} \mathcal{C} \) and a col-
lection of arrows \( \text{Mor} \mathcal{C} \) between these objects, together with a composition operation \( \circ \)
on arrows such that:

\begin{enumerate}
\item For every \( f : a \to b \) and \( g : b \to c \) there is an arrow \( g \circ f : a \to c \).
\item For every object \( a \) in \( \text{Ob} \mathcal{C} \) there is an arrow \( 1_a : a \to a \) (called the “identity
   arrow”) such that \( 1_a \circ f = f \) and \( f \circ 1_b = f \) for every arrow \( f : a \to b \).
\item \( \circ \) is associative.
\end{enumerate}

\textsuperscript{23}Many have been proposed, most notably in [14].
Within a category we then say that a given arrow $f: a \to b$ is an isomorphism if there is another arrow $g: b \to a$ going in the other direction such that $g \circ f = 1_a$ and $f \circ g = 1_b$.

**Definition 4.4.2.** A groupoid is a category in which every arrow is an isomorphism.$^{24}$

To be fully precise, I will refer to the sought-after entity as a next-level set, to make clear that what is sought is the notion of a set exactly one level up. Now, as suggested above, there are two plausible answers to the question “What is a next-level set?”:

(C) A next-level set is a category

(G) A next-level set is a groupoid

How can we approach (C) and (G) as philosophical theses? A reasonable method is to state a general principle defining “higher-level”-ness and then show that this principle, when specialized to sets, yields the desired answer. Consequently, choosing between (C) and (G) reduces to choosing between these two general principles. The issue, as we shall see, cannot be settled as cleanly as that – but it is with this methodological optimism that I will begin.

In support of (C) we will consider the principle that Makkai [113] has called the *Fregean Imperative*:

(FI) A next-level $X$ is the type of thing formed by the collection of all $X$s

$^{24}$It is very important to note that a groupoid does not have to be defined in this manner (even if this is the most concise way to do so). The notion of a groupoid actually pre-dates the notion of a category and was initially seen as a generalization of the notion of a group. In other words, we are by no means forced to define a groupoid as a category with an extra property.
In support of (G) we will consider a principle that I will call *Bishop’s Imperative*:25

(BI) A next-level $X$ is the type of thing between each of whose constituents there are (at most) $X$-many identifications

The phrasing of (BI) at this level of generality is cumbersome, so let me try to clarify further. Suppose $X$ is a collection of things. Then the next-level $X$ is a collection of things such that for any two such things, the collection of ways in which they can be identified is itself a collection of type $X$. Consider the following simple but crucial example. Let $X$ be a truth-value, i.e. $X$ is either *true* or *false*.26 Then by (BI) a next-level truth value (call it $Y$) is the type of thing between each of whose constituents there are (at most) truth value-many identifications (where “truth value-many” simply means that between each pair of constituents of $Y$ there either is (“true”) or isn’t (“false”) an identification). Namely, $Y$ is a collection of things each of which is uniquely separated from all the others: whether a thing in $Y$ is the same as another thing in $Y$ is answerable merely by a “yes” or a “no”. In short, next-level truth values are collections of things, uniquely determined by their elements. Such entities have been known to operate under the name of “sets”. Thus, according to (BI), sets are next-level truth values.

So what – if anything – can we say about (BI) and (FI)? There are two questions we must tackle:

(Q1) Does (C) (respectively (G)) follow from (FI) (respectively (BI))?  

25The name is a reference to E. Bishop’s view on sets as can be found e.g. in [22,143]. But there are also obvious connections with the behaviour of identity types in (intensional) MLTT [124].

26One can thus think of $X$ as an element of the set {⊤,⊥} or {0,1}. But it is in no way necessary to think set-theoretically.
Are there any philosophical reasons to prefer either (BI) or (FI)?

I will take these questions in turn.

4.4.1. Does (G) follow from (BI)?. Voevodsky [180] has put forward an argument in support of (G).\textsuperscript{27} It goes as follows:

(1) **Categories** can be thought of as **groupoids** with extra structure. In particular, if $G = \text{Ob} \mathcal{C}$ is a groupoid then $\text{Mor} \mathcal{C}$ can be thought of as a function

$$\text{Mor} \mathcal{C} : G \times G \to \text{Set}$$

picking out a set of arrows between any two objects of $G$ satisfying the appropriate properties.\textsuperscript{28}

(2) **Partially ordered sets** can be thought of as **sets** with extra structure. In particular, if $(X, \leq)$ is a partially ordered set this means that for any $x, y \in X$ we have that either $x \leq y$ or not. Thus, the ordering $\leq$ on $X$ can once again be thought of as a function

$$\leq : X \times X \to \{0, 1\}$$

satisfying the appropriate properties.

(3) If categories are groupoids with extra structure then lower-level categories are partially ordered sets.

\textsuperscript{27}It might be unfair to call it an “argument” since it was not presented as one, at least in the philosophical sense of the term. Nevertheless, the kind of thinking that this “argument” represents was crucial in the development of UF. To see this, consider what Voevodsky says (also in [180]): “One of the things that made the “categories” versus “groupoid” choice so difficult for me is that I remember it being emphasized by people I learned mathematics from that the great Grothendieck in his wisdom broke with the old-schoolers and insisted on the importance of considering all morphisms and not only isomorphisms [...] and that this was one of the things that made his approach to algebraic geometry so successful.”

\textsuperscript{28}The way I’ve laid out Definitions 1 and 2 above should make it clear that this is not the usual way of thinking about categories. Formally, however, these two ways are equivalent (in an appropriate sense).
Therefore, next-level sets are not categories. Instead, categories are higher-level posets.

As an argument purely in support of (G), this is clearly hopeless since the analogy drawn between (1) and (2) presupposes (G). For in saying that (1) is the higher-level analogue of (2) one must already accept that groupoids are the higher-level analogues of sets. If anything, what this argument shows (successfully) is that if one takes next-level sets to be groupoids then one can view categories as next-level partially ordered sets. But this is not something that we care about here, nor is it something that, I think, anyone would regard as particularly controversial. Furthermore, this argument in no way relies on (BI) or on any other general principle about “higher-level”-ness. As such, Voevodsky’s argument in itself cannot provide any justification for the thesis that from (BI) it follows that next-level sets are groupoids. (Nevertheless, the heuristic that categories are next-level partially ordered sets is a very useful one.)

So let us instead try to get to (G) from (BI) in a more direct manner. Let $X$ be a set. As summarized above, $X$ is a collection of things (elements) such that for any two elements there is only one unique way in which they can be identified. Put differently: for any two elements of a set we care only about whether they are identical and not about in how many ways they can be regarded as identical. Thus, according to Bishop’s imperative, a next-level set would be a collection of things such that there are “set-many” ways in which any two such things can be identified.

Such a collection is now starting to sound very much like a groupoid if we think of the “set-many” identifications between any two objects as the set of isomorphisms between them (in a groupoid, recall, every arrow is an isomorphism but there is no requirement that there should be a unique isomorphism between any two objects: there could be “set-many” of them). The remaining difficulty is whether or not it makes sense to regard
“isomorphisms” as “identifications”. If it does, then we are done: (BI) demonstrates that a next-level set is a groupoid. But what licenses us to say that isomorphism is the same thing as identity? Surely we can regard isomorphic objects as identical if we so wish, but that does not mean that isomorphism is the same thing as identity.

Unless we have some formal principle that allows us to regard isomorphism as identical to identity, then we cannot advance any further towards (G). The axiom of univalence (UA) provides just such a formal principle – but of course we cannot invoke it here, while engaged in “pure” philosophy. For even though it is straightforward to see that (BI)+(UA) ⇒ (G), I am interested in the much stronger statement (BI) ⇒ (G). And in arguing for this stronger statement, it is not immediately clear to me how to overcome this difficulty.

4.4.2. Does (C) follow from (FI)?. On the other hand, with his acceptance of the Fregean imperative in the background, Makkai supplies an argument in favour of (C). It goes as follows:

1. The identity relation between (abstract) sets can only be applied to their “elements” – in fact this characterizes (abstract) sets.

2. Sets are not “elements” of other sets.

3. You cannot ask of any two sets whether they are identical.

4. Therefore the collection Set of all sets is not a set, since you cannot ask of any of its elements if they are equal.

29If one trusts that (BI)+(isomorphism=identity) leads to (G) then the only thing left is to see (UA) as an axiom asserting (isomorphism=identity). This is argued for, convincingly, by Awodey in [9]. It needs to be clarified, however, that (UA) does not collapse isomorphism to identity by coarsening the latter. Rather it expands identity to isomorphism by defining it as homotopy equivalence. In particular, it is not that there were things we could separate before asserting (UA) that we no longer can after asserting it – (UA) does not coarsen identity; it defines it.

30Note that “sets” here means “abstract sets” i.e. the kind of entities described e.g. by the FOLDS theory of sets in [113]. For further discussion on the distinction between “abstract” and “concrete” in this context as well as more discussion on Makkai’s argument see [122].
(5) Instead you can define what it means to have an isomorphism of sets, but that requires having a category structure on $\textbf{Set}$.

(6) Therefore, next-level sets are categories.

The crux of the argument is in (4) where it is claimed that the kind of structure that the collection of all sets forms is the structure that allows us to formulate a criterion of identity for them, viz. isomorphism. This is because the collection $\textbf{Set}$ of all sets is not itself a set and since it is only possible to ask of elements of sets whether they are identical it is impossible (!) to ask of sets themselves whether they are identical. What we can do instead is ask whether they are isomorphic – but that requires having a category structure on $\textbf{Set}$! Therefore, the collection of all sets is a category and as such, by (FI), next-level sets are categories.$^{31,32}$

Ingenious as this argument may appear, there is clearly a hidden premise operating here, namely that the structure borne by the collection of all things of type $X$ is the kind of structure that allows one to establish/define a criterion of identity for things of type $X$. In the case of sets, this hidden premise specializes to the following: the structure borne by the collection $\textbf{Set}$ of all abstract sets is the kind of structure that allows one to

---

$^{31}$A very similar way of separating sets from “categories” can also be found in the work of Martin-Löf:

A category is defined by explaining what an object of the category is and when two such objects are equal. A category need not be a set, since we can grasp what it means to be an object of a given category even without exhaustive rules for forming its objects. [...] To define a category it is not necessary to prescribe how its objects are formed, but just to grasp what an arbitrary object of the category is. ([124], pp. 21-22)

In the above quote “category” does not necessarily refer to “category” in the sense of category theory, but rather to the kind of thing signified by the symbol $\textbf{Type}$ in type theory, i.e. the collection of all types. If type theory contains the appropriate type-formers and axioms (e.g. $\Pi$-types with the usual rules) then $\textbf{Type}$ will also carry the structure of a category-in-the-sense-of-category-theory. But – for Martin-Löf – the structure of a category-in-the-sense-of-category-theory does not immediately “emerge” from the above-sketched way of separating “categories” from sets. (It is also worth noting that Martin-Löf uses “sets” here for what we would now call “types” in MLTT.)

$^{32}$Very similar arguments have been put forward by several authors arguing in favour of some variant of category-theoretic structuralism in the philosophy of mathematics (e.g. [6,89] and also [95]). The main idea there is to separate the way in which an object is “given” in category theory (schematically, abstractly, up to isomorphism) from the way in which an object is “given” in set theory (by being constructed directly out of simpler objects).
define/establish a criterion of identity for sets. This criterion of identity, Makkai assumes, is isomorphism and, Makkai claims, it is only in the setting of categories that we get an abstract definition of isomorphism between two objects (along the lines indicated after Definition 4.4.1).

But then we are faced with the following dilemma. If one is to understand the argument as claiming that the definition of what it is for something to be the next-level $X$ is to be that kind of structure that provides the collection of all things of type $X$ with a criterion of identity, then surely this definition is not specific enough to pick out categories as next-level sets. Indeed this description is perfectly compatible with taking groupoids to be next-level sets, since really all that is needed in order to provide individual sets with a criterion of identity is the underlying groupoid of $\textbf{Set}$. After all, if we only care about providing a criterion of identity then what do we need non-invertible arrows for at all?

On the other hand, if the argument is to be understood as saying that it is because a category structure provides a criterion of identity between elements of $\textbf{Set}$ that $\textbf{Set}$ is to be regarded as a category, then the argument becomes viciously circular. For in order to conclude that $\textbf{Set}$ is a category one has to assume that it is one. More precisely, the argument would go as follows:

(1) If $\textbf{Set}$ is a category then it provides a criterion of identity for sets

(2) If something provides a criterion of identity for $\textbf{Set}$ then it is a type of structure one level higher than sets.

(3) Therefore $\textbf{Set}$ is a category

In order to conclude (3) from the two conditionals (1) and (2) one must assert that $\textbf{Set}$ is a category – but this is the desired conclusion.
Neither Makkai’s nor Voevodsky’s arguments, I conclude, provide satisfying positive answers to \((Q_1)\). In failing to do so, they illustrate for us the main difficulty of approaching the question through an investigation of (FI) and (BI): both principles are equally compatible with giving the opposite of the desired answer. More precisely, (FI) is compatible with (G) because \(\text{Set}\) also bears the structure of a groupoid\(^{33}\) and (BI) is compatible with (C) since any groupoid can be regarded as category with an extra property (all arrows are invertible).

This leads us to the following crucial question: are groupoids to be regarded as categories with an extra *property* (that all arrows are invertible) or are categories to be regarded as groupoids with extra *structure* (non-invertible arrows between objects)? Clearly, this is at root a question about fundamentality. We are asking: which structure is more fundamental, groupoids or categories?

If we take this line, then the acceptance of (FI) or (BI) should not really be seen as directly justifying that next-level sets are categories or groupoids. It should instead be understood as a justification for the *fundamentality* of categories over groupoids or groupoids over categories. On this view (FI) (respectively (BI)) implies that categories (respectively groupoids) are somehow more fundamental than groupoids (respectively categories).

But what exactly is “fundamentality” even supposed to mean here? It seems hopeless to give it a precise *technical* meaning (e.g. by saying that \(X\) is more fundamental than \(Y\) if \(Y\) is the same type of thing as \(X\) but with an extra property) without deciding the question in advance. Furthermore, any such technical criterion will depend on some kind of fixed background theory (e.g. set theory) in which both groupoids and categories

\(^{33}\)Simply take the underlying groupoid (or “core”) of the category of sets and functions.
will already have been defined (e.g. as structured sets). But choosing such a background
theory seems to make the question meaningless, since the notions we want to compare are
now defined in terms of something even more fundamental. For example, if we define both
groups and topological spaces as structured sets does it make sense to ask which is more
fundamental? Or even to try and come up with a technical set-theoretic criterion to make
the question precise? It seems equally hopeless to try to understand “fundamentality”
in terms of historical priority, viz. by figuring out whether groupoids or categories arose
first in mathematical practice.\footnote{For the interested reader, groupoids first appeared in the work of Brandt \[24\] long before categories
were defined. According to Ronnie Brown, there is apparently a rumour that Mac Lane and Eilenberg
(the inventors of category theory) were inspired by groupoids in their definition of a category although
this has been denied by Eilenberg (cf. \[25\], p. 118).
}

On the other hand, if we are going to rely on a non-technical and non-historical notion
of fundamentality this would seem to depend on having some kind of pre-formal grasp
of categories or groupoids, analogous to how in the case of sets we have some pre-formal
grasp of what it means for something to be a collection of things. Moreover, since both
(BI) and (FI) can plausibly be read as pre-formal principles it makes more sense to see
them as being applicable to pre-formal notions.\footnote{At least in the case of categories, several arguments have been proposed over the years against the
possibility of such a pre-formal account that does not on some level depend on some (naive/pre-formal)
set theory. Feferman \[41\] argued that the notion of “collections and operations” is required in order
to talk about categories and so any pre-formal talk of categories assumes some pre-formal set theory.
The debate continues all the way to the present day: see \[42, 90, 122, 135\] for some of the most recent
installments. My own attitude to such objections is that the notions of collection and operation pre-dated
by some margin even the earliest formal use of “abstract” sets in mathematics. Therefore, to rely on
concepts such as “collection”, “thing” and “operation” when laying out a pre-formal account of anything
(including sets) should be granted as the bare minimum if any kind of philosophy of mathematics is to
take place at all.}

So I will now go on to develop a pre-formal account of both categories and groupoids.
I will then argue that with this pre-formal account in hand, it is indeed true that (BI)
implies that (pre-formal) groupoids should be regarded as next-level (pre-formal) sets and
similarly for (FI). As is clear, to do so I will rely on a notion of a pre-formal set. By this
I will mean simply the naïve idea of a set as a collection of things. These “things” ought
to be thought of as “structureless” in a naive sense: there is nothing we can say about them other than the fact that they are parts of the collection at hand. The pre-formal notion of set I have in mind, then, is that of a “purely extensional” collection: an abstract set.36

4.5.1. Pre-formal Groupoids. The basic idea is this: a groupoid is a shape. As with any shape, it has edges and it has vertices although we allow there to be multiple edges between any two vertices. Importantly, the edges are not directed. Intuitively this means that if you can walk from one vertex to another along an edge, then you can just as easily walk back along the same edge. In short, a groupoid is a collection of things (which we can regard as points) some of which are connected by a collection of edges (which we regard as lines or paths). A pre-formal groupoid thus looks something like this:

\[\begin{array}{c}
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\bullet \\
\end{array}\]

I think it is fairly intuitive that pictures like the above represent objects intuitively comprehensible as shapes and I will not argue for it any further. This is all my notion of a pre-formal groupoid consists in: a graph with undirected edges such that there can be multiple edges between any two nodes. This captures enough of the intuition that groupoids are mathematical structures encoding the homotopy types of (not necessarily path-connected) topological spaces. Although groupoids are not used exclusively for these

36I take it that this concept is intuitive enough to require no further analysis. It is a different issue whether this notion of an abstract set is compatible with the standard Zermelian conception of sets, but I will say no more about this here. I am here interested in a comparison of categories and groupoids modulo some fixed common conception of set, and changing this conception of sets does not affect the comparison being made, nor any of the conclusions I reach. My account could easily be modified to accommodate a more (or even less) sophisticated understanding of a pre-formal set, since what I care about here is to compare (C) and (G). Insofar as they depend on the same account of pre-formal sets (whatever that may be), the arguments below can be run just the same.
purposes, it is still fair to say that most of their uses in one way or another reflect this intuition. And it is only this intuition on which we will rely.\textsuperscript{37}

It is also important to note that this pre-formal notion leaves out a crucial part of the formal definition of groupoids, namely the existence of identities and composites together with coherence conditions dictating how they interact. I justify this omission for three reasons: firstly, because it is to some extent implicit (pre-formally) in the notion of a shape; secondly, because these algebraic conditions are not essential to the points I make below; thirdly, because in the case of 1-groupoids these “coherence conditions” are very few and can easily be written down.

4.5.2. Pre-formal Categories. The immediate temptation is to take categories to be shapes just like above but also with a notion of direction in their edges. But the situation is more complicated. (Moreover, to proceed like this would also have been to decide the question of fundamentality: for certainly then groupoids would come out more fundamental than categories in the same way that manifolds can be seen as more fundamental than oriented manifolds.)

In the case of groupoids there is clear spatial aspect that motivates their definition, which is also why it is reasonable to call them “shapes”. In particular, it is faithful to the way groupoids are being used to regard the dots as points and the lines as (undirected) paths. On the other hand, the motivation for the definition of categories had mainly to do with the organization of algebraic and topological structures into structured collections.\textsuperscript{38}

Therefore, categories are best thought of, pre-formally, as collections of structures and not as directed graphs. A particular feature of a “structure” is that it is exemplified by

\textsuperscript{37}For more on the applicability of groupoids see \cite{25} and for a more detailed elaboration of a similar point of view of pre-formal groupoids as “forms” see \cite{121}.

\textsuperscript{38}This was done in order to formalize certain constructions that were becoming ubiquitous in algebraic topology at the time. In particular, to formalize the idea that algebraic information could be attached to topological spaces in a natural or compatible manner. As Mac Lane is said to have remarked, category theory was invented not in order to study functors, but in order to study natural transformations.
many (concrete) instantiations. To recognize that the same structure is being instantiated by two distinct objects is to recognize that these two objects are related in at least one way; namely, that they can be seen to bear a common structure. Every such way of relating these two objects we can represent with an arrow. When structures are represented by points and ways of relating them by arrows then we get the usual picture of a category as something that looks like this:

\[ \text{\includegraphics[width=0.5\textwidth]{category_diagram.png}} \]

But the above picture should not, I urge, be viewed as a graph in the usual mathematical sense of a set of vertices and a set of edges etc. Formally, of course, one can do so. But the pre-formal understanding of categories I have just sketched goes beyond the pre-formal understanding of a “directed graph”. Rather, a pre-formal category in my sense is a “directed graph” in which nodes are to be understood as structures and arrows are to be understood as (witnesses of) structure-similarity between these structures. A pre-formal category is therefore a collection of structures together with a network of structural similarities between them.

This discussion reveals an oft-neglected fact, which is that categories lead a double life. They can be understood, firstly, as mathematical structures or, secondly, as collections of mathematical structures. More precisely, categories can be regarded as algebraic/combinatorial objects in their own right (transitive reflexive directed graphs with extra properties) or as formalizations of collections of such objects (e.g. Set, Grp). The first aspect has proved extremely useful in algebraic topology, e.g. to encode the essentially combinatorial nature of certain topological constructions like CW-complexes.
The second aspect is historically also related to algebraic topology but has found applicability in many other fields and is used primarily for organizational reasons. We may call the former the spatial aspect of categories and the latter the organizational aspect.\textsuperscript{39} In those terms, the notion of pre-formal categories outlined above corresponds to the organizational aspect.

This dual aspect also clearly makes sense in the case of groupoids: there are spatial and organizational notions of a groupoid, understood in an exactly analogous manner. The distinction I have drawn between pre-formal groupoids and pre-formal categories can thus be summarized as follows: my pre-formal groupoids are spatial whereas my pre-formal categories are organizational. This difference will be essential in the argument presented below.

4.5.3. Bishop or Frege? We can now revisit (Q\textsubscript{1}): does (FI)/(BI) imply that next-level sets are categories/groupoids? Let me start with (FI). A next-level set, according to (FI), is the kind of thing exemplified by the collection of all sets. This collection of all sets consists of objects with a certain “purely extensional” structure, but they are certainly structures of some sort. And the way we have of relating these purely extensional structures is by pairing up their elements, viz. by defining functions between them. And this, indeed, gives rise exactly to a pre-formal category as outlined above.

On the other hand, according to (BI), a next-level set is the kind of thing between each of whose constituents there is (at most) a set of identifications. This means that a next-level set is a collection of nodes together with, for any pair of nodes, a (possibly empty) set of “identifications”. If we take these “identifications” to be structureless “edges” we get exactly the picture of the pre-formal groupoid I outlined above. Namely, between any two nodes there will be a certain number of lines connecting them. As such, both

\textsuperscript{39}To clarify, my “spatial categories” are not to be confused with the study of “categories of (generalized) spaces” in the vein of e.g. [74].
(FI) and (BI) render support to (C) and (G) when the terms “category” and “groupoid” are understood in the pre-formal way I have indicated: according to (FI) next-level sets are (organizational) categories whereas according to (BI) next-level sets are (spatial) groupoids.

We are now ready to shift our attention to (Q2): what can we say about (FI) and (BI) as self-standing principles? To my mind, (FI) is susceptible to one main objection: the collection of things of type X usually bears many natural (non-equivalent) structures. For example, as argued above, the collection of all sets does not just bear the structure of a category with arrows given by the functions between sets. Why should the structure of Set somehow uniquely emerge when we collect all sets together? And even if it did why should we choose functions as the arrows? There are alternatives after all: for instance we have the category Rel with objects the same as Set but with arrows all the (not necessarily functional) relations between sets. On the whole, (FI) seems to be too “loose” a guiding principle.

As for (BI), the main difficulty is how to understand the passage from the “empty” collection to the “singleton” collection. Recall that, according to (BI), a lower-level set is a truth value. So what then is a lower-level truth value? Let us once again reverse-engineer:

\[(\ast) \quad \text{A lower-level truth value is the type of thing } Y \text{ such that there can be at most } Y \text{-many identifications between elements of a truth value.}\]

The phrase “elements of a truth value” raises the obvious question: how can the empty (or “false”) truth value have an element? The answer is that it cannot. The statement of (\ast) only makes sense for the inhabited (or “true”) truth value. And there the answer is clear: there can only be a single unique identification between the single element of the inhabited truth-value. As such, a lower-level truth value is a singleton. Thus, according to
(BI), the notion of a singleton is more fundamental (in the sense of being of a lower level) than the empty collection. Put more provocatively, but less accurately: non-existence is a higher-level form of existence.\textsuperscript{40}

What in the world are we talking about? Have we bumped our heads against the limits of our language? Yes, we have. We are translating into natural language constructions that only properly make sense in formal systems. We cannot expect the sense of these constructions to survive that transition. But neither should we surrender to uninterpreted formalism. So what I’ll say is that such considerations as the above might prove decisive for someone holding a particular metaphysical view (such a person will also, of course, need to have a pre-formal account of all the terms being used, i.e. “singleton”, “collection”, “truth value” etc.). For example, a so-called compositional nihilist might not want to countenance (FI) purely because it considers $\text{Set}$ as a thing over and above its individual constituents (i.e. sets). Someone else might think that it is absurd to give ontological priority to the singleton set over the empty set as (BI) seems to force us to do. But absent some such metaphysical thesis in the background, I don’t see any philosophically decisive reason to pick either (BI) or (FI). We have reached an impasse.

4.5.4. Next-level sets are groupoids. As I’ve argued in Section 4.4 the notion of “higher-level” requires some level of mathematical precisification if it is to prove at all tractable to philosophical investigation. Similarly, in order to answer ($Q_2$) in any interesting way, we must embed it in a concrete mathematical context. This will provide a way out of the impasse reached in the previous subsection. The most relevant context here is the foundations of mathematics. And since (FI) and (BI) are really best understood as mathematical justifications for (C) and (G) respectively the better question to ask (given

\textsuperscript{40}If this discussion is straining the reader’s credulity allow me to clarify that what I am doing here is informally describing the hierarchy of $n$-types in as formalized in (a given) homotopy type theory. What is here expressed in prose form has been fully formalized, cf. [171], Chapter 7.
what has been said so far) is this:

(Q$_3$) Should we pick spatial groupoids or organizational categories 
(and the hierarchies they generate) as the basic objects of our 
(structuralist) foundations?

And here, finally, I believe a clear answer emerges: we should prefer spatial groupoids 
over organizational categories.

The decisive objection against organizational categories as the building block in a 
hierarchy of basic objects for a foundation of mathematics is that, as the name suggests, 
the very notion of an organizational category pre-supposes that it organizes something. 
So if an organizational category is to be regarded as a basic structure of a foundation, 
then the basic structures of that foundation will themselves be structures organizing 
other structures. This is like saying that the foundations of a house are not the bricks 
that go into its construction but the cardboard boxes in which these bricks are stacked 
and transported. If categories are to be fruitfully regarded as the basic structures of a 
foundational system then it cannot be merely by virtue of the fact that they provide a way 
to organize other, previously constructed, structures. The basic objects of any foundations 
should simply be the basic structures that one can then organize into whatever shape or 
form one wishes.

Thus, in a hierarchy of basic objects, the lower-level objects should be understandable 
as degenerate versions of the higher-level ones and not as entirely distinct kinds of things 
(as distinct as, say, bricks are from the cardboard boxes that store them). Does it really 
make sense to say that sets are degenerate (discrete) categories because the collection of 
all sets forms a category? As far as I am concerned it does not: bricks are not degenerate
versions of cardboard boxes. On the other hand, it seems to me far more intuitive to say that a set is a degenerate (spatial) groupoid. For in this case we are talking about a degenerate shape and there has hardly ever been a problem understanding parts of shapes as degenerate versions of more complicated ones, e.g. lines as degenerate triangles.

Thus, if categories are to be regarded as the basic building blocks of a (hierarchy of) basic objects, it must be in their spatial aspect, i.e. by being regarded as (transitive, reflexive) directed graphs with an associative composition operation. But there are now two serious problems. Firstly, (FI) no longer applies: if we were to apply (FI) to sets we would not end up with spatial categories as next-level sets. For why would a transitive, reflexive, directed graph be the type of thing we get when we gather all sets together?

Secondly, if we are indeed to view categories as directed graphs then these clearly seem to be less fundamental than spatial groupoids since the latter can roughly be understood as undirected graphs. So if we are to compare spatial categories and spatial groupoids with respect to their suitability to be the next-level sets in a foundational hierarchy then it would seem much more natural to regard categories as groupoids with extra structure (namely a *non-invertible* direction). But then building a foundation whose basic building blocks are spatial categories and then defining spatial groupoids by adding more properties, seems analogous to axiomatizing the universe of partially ordered sets (i.e. defining something like a “ZF poset theory”) in order to then define sets within it as those posets with a trivial ordering.\footnote{Of course this is not to say that there is no way that there could be *independent* spatial motivation for categories. Indeed, such a perspective is outlined by Grandis in [52]. What I am saying, however, is that such “directed” perspectives are clearly less fundamental than undirected ones – and therefore clearly less suitable for a foundation of mathematics.}

This is where Voevodsky’s aforementioned “argument” helps clarify the situation: in traditional set-theoretic foundations it would seem absurd to try and axiomatize the “cumulative hierarchy of partially-ordered sets”. In the realm of structuralist foundations,
it should strike us as equally absurd to try and axiomatize the “hierarchy of higher categories”. We should, instead, try to axiomatize the hierarchy of higher groupoids, i.e. the hierarchy of homotopy types. And then higher categories should be studied as structures defined on those higher groupoids, just like partially ordered sets are studied as structures on sets.

The answer to (Q₃), I conclude, is that we should prefer spatial groupoids over organizational categories as next-level sets in the context of the foundations of mathematics. This would be an entirely uninteresting conclusion if there weren’t any foundations of mathematics that took the question of next-level sets as a serious (indeed axiomatic) design constraint. Since I have argued on purely philosophical grounds that the correct notion of a higher-level set is a groupoid, not a category, everything I have said in this Chapter now provides a decisive reason to pick one over the other.

It should be made clear that there are in addition strong pragmatic considerations that corroborate with this choice. In particular, intensional Martin-Löf Type Theory provides a remarkably simple description of the structure of homotopy types via the four rules for identity types. No such description is currently available for directed homotopy types, i.e. types that natively bear the structure of ∞-categories rather than ∞-groupoids. The availability of a very well-understood and well-developed formal system whose basic objects bear the structure of (higher) groupoids is therefore a very good pragmatic reason to take groupoids as higher-level sets, even without the kind of philosophical justification that I have endeavored to provide.

Nevertheless, I believe that the significance of the idea of a higher-level set being a groupoid is not limited to the rather narrow domain of structuralist foundations of mathematics. It is relevant, I submit, to far broader philosophical concerns because it provides philosophy with an entirely new notion of a higher-level collection of things to
consider. Among other things, this allows us to re-visit one of the core ideas of logic since antiquity: that to define a property (or concept) is to define a feature common to a multitude of objects.\footnote{Whether one wants to be a realist, anti-realist or whatever else about the nature of these objects does not change the particular definition of “concept”, only what we mean by “object”.} A higher-level property is then a feature common to a multitude of properties and so on. But what if a higher-level property were simply one in which we tracked the ways in which collections of objects share a feature, rather than merely assert that they do? Then we get, essentially, a notion of a higher-level property being a shape of some sort. Groupoids (possibly higher ones) in UF provide a formalization of this idea. It is imperative that the philosophical implications of this formal development be explored. To that end, in the next Chapter I will develop a mathematical logic with which this can be done.

Appendix: Homotopy equivalence as FOLDS equivalence

In this appendix we will prove an observation of Makkai that FOLDS equivalence coincides with homotopy equivalence in the case of semi-simplicial sets. We recall the basic notions, following [70, 110] in our presentation. Let $\Delta$ denote the category whose objects are the ordered sets $[n] = \{0 \leq 1 \leq \ldots \leq n\}$ for each $n \in \mathbb{N}$ and arrows the monotone functions between them. We let $\Delta_+$ denote the subcategory of $\Delta$ consisting of the same objects but having as arrows only injective functions. We write $i: \Delta_+^{\text{op}} \to \Delta^{\text{op}}$ for the canonical inclusion. A simplicial set is a functor $\Delta^{\text{op}} \to \text{Set}$ and a semi-simplicial set is a functor $\Delta_+^{\text{op}} \to \text{Set}$. We write $\text{sSet}$ (resp. $\text{ssSet}$) for the category $\text{Set}^{\Delta^{\text{op}}}$ (resp. $\text{Set}^{\Delta_+^{\text{op}}}$) of simplicial (resp. semi-simplicial) sets and we will write $i^*: \text{sSets} \to \text{ssSets}$ for the obvious forgetful functor induced by $i$. As is standard, we write $\Delta[n]$ (resp. $\Delta_+[n]$) for the $n$-simplex (resp. semi-simplex), $\partial \Delta[n]$ (resp. $\partial \Delta_+[n]$) for the boundary of the
An $(n$-(semi-)simplex and $\Lambda^i_{+}[k])$ for the $i^{th}$ horn (resp. semihorn) of the $n$-(semi-)simplex. By (semi-simplicial) boundary inclusions we will mean all inclusions of the form $\partial_n: \partial[n] \rightarrow \Delta[n]$ (resp. $\partial^+_n: \partial[n] \rightarrow \Delta[n]$) and write $\mathcal{I}$ (resp. $\mathcal{I}_+$) for the set of all of them. A Kan complex is a simplicial set $S$ that satisfies the Kan condition: for all $k$ and all $0 \leq i \leq k$ and for any arrow $\lambda: \Lambda^i[k] \rightarrow S$ there exists an arrow $\delta: \Delta[k] \rightarrow S$ such that $\lambda = \delta \circ h_{i,k}$. We will also call a semi-simplicial set Kan when it is the image of a Kan complex under $i^*$. We recall that the standard model structure ([70], Definition 3.2.1) on $sSets$ takes its weak equivalences to be the maps whose geometric realization induces a weak homotopy equivalence of spaces. We say that two Kan complexes are homotopy equivalent if their realizations are homotopy equivalent as spaces. All other homotopical notions we use below will refer to the standard model structure.

A FOLDS-signature is a finite, one-way and reverse well-founded category $\mathcal{L}$. These latter two conditions are equivalent to requiring the existence of a function $l: \text{ob}\mathcal{L} \rightarrow \mathbb{N}$ such that $l(K) = 0$ for some $K \in \text{ob}\mathcal{L}$ and for all non-identity arrows $f: K \rightarrow K'$, $l(K) > l(K')$. An $\mathcal{L}$-structure is a functor $M: \mathcal{L} \rightarrow \text{Set}$ and a homomorphism of $\mathcal{L}$-structures is a natural transformation between the corresponding functors. We write $\text{Str}(\mathcal{L})$ for the functor category $\text{Set}^\mathcal{L}$, $y$ for the Yoneda embedding and $\hat{y}$ for subfunctor of $y$ that misses the identity on the given object, i.e. $\hat{y}X(X) = yX(X) \setminus \{1_X\}$. For $K \in \text{ob}\mathcal{L}$ a $K$-boundary of an $\mathcal{L}$-structure $M$ is a natural transformation $\delta: \hat{y}K \Rightarrow M$ and we define the fiber of $M(K)$ over $\delta$ as the set $M(K)[\delta] = \{a \in M(K): \partial a = \delta\}$. Now let $M, N$ be $\mathcal{L}$-structures and let $f: M \rightarrow N$ be an $\mathcal{L}$-structure homomorphism and let $\delta$ be a $K$-boundary of $M$. This induces a $K$-boundary $f \circ \delta$ of $N$ by composition with $f$ and thereby induces a map $f_\delta$ on the fibers that takes $a \mapsto f_K(a)$. (The fact that this is well-defined follows from the naturality of $f$.) We now say that $f$ is fiberwise surjective if and only if for all $K$ in $\mathcal{L}$ and all $K$-boundaries $\delta$ the map $m_\delta: M(K)[\delta] \rightarrow M(K)[m \circ \delta]$
is surjective. We say that $M$ and $N$ are $FOLDS \ L$-equivalent and write $M \simeq_L N$ if there exists an $L$-structure $P$ and fiberwise surjective homomorphisms $m: P \to M$ and $n: P \to N$. We may immediately observe that $\Delta_+^{op}$ is a FOLDS signature since it is clearly both well-founded and each of its objects, being a finite set, is the codomain of finitely many maps. Therefore all FOLDS notions developed above make sense relative to $\Delta_+^{op}$ and this allows us to state our main theorem which is based on an observation of Makkai in [112].

**Theorem 4.5.1.** Two Kan complexes $S$ and $T$ are homotopy equivalent if and only if $i^*S \simeq_{\Delta_+^{op}} i^*T$.

Recall that in any category a map $f: C \to D$ has the right lifting property (RLP) with respect to a map $g: A \to B$ if for every $h: A \to C$ and $k: B \to D$ such that $fh = gk$ there exists $d: B \to C$ that makes the two obvious triangles commute.

**Lemma 4.5.2.** Take any FOLDS signature $\mathcal{L}$ and a homomorphism of $\mathcal{L}$-structures $f: M \to N$. Then $\eta$ is fiberwise surjective if and only if $\eta$ has the right lifting property with respect to the inclusion $\hat{y}K \hookrightarrow yK$ for all $K \in \mathcal{L}$.

**Proof.** Necessity is immediate from the definitions. For sufficiency, let $f: M \to N$ have the right lifting property with respect to all inclusions $\iota_K: \hat{y}K \hookrightarrow yK$ and let $\alpha$ be a $K$-boundary of $M$, i.e. a natural transformation $\alpha: \hat{y}K \Rightarrow M$. It suffices to show that the induced map $f_\alpha: M(K)[\alpha] \to N(K)[f \circ \alpha]$ is surjective. So take any $b \in N(K)[f \circ \alpha]$. By Yoneda this can be identified with $b: yK \Rightarrow N$ such that $\partial b = b \circ \iota_K = \eta \circ \alpha$. By assumption there exists a lift $\lambda: yK \Rightarrow M$ such that $f \circ \lambda = b$ and $\lambda \circ \iota_K = \alpha$. By Yoneda this means that there is an element $\lambda \in M(K)$ such that $f_\lambda = b$. Also since $\lambda \circ \iota_K = \partial \lambda = \alpha$ this means that $\lambda \in M(K)[\alpha]$. Therefore $f_\alpha(\lambda) = b$ and we are done. □
Note that in the case of $\text{sSet}$, the inclusions of the form $\hat{y}K \hookrightarrow yK$ are the boundary inclusions $I$ and in the case of $\text{ssSet}$ they are the semi-simplicial boundary inclusions $I_+$. The forgetful functor $i^*: \text{sSets} \to \text{ssSets}$ has left and right adjoints $i_! \dashv i^* \dashv i_*$ obtained as left and right Kan extensions. These adjunctions allows us to transfer to model structure of $\text{sSets}$ to $\text{ssSets}$ (cf. [173]) and in particular allow us to talk about weak equivalences between semi-simplicial sets. We now require two lemmas.

**Lemma 4.5.3.** For any $n \geq 0$, we have $i_! \Delta_+[n] \cong \Delta[n]$ and $i_! \partial \Delta_+[n] \cong \partial \Delta[n]$. In particular, we have that $i_! I_+ = I$.

**Proof.** Since $i_! \Delta_+[n]$ is the left Kan extension of a representable functor we have $i_! \Delta_+[n] \cong \Delta(-, i([n])) = \Delta[n]$. Exactly analogously we get $i_! \partial \Delta_+[n] \cong \partial \Delta(-, i([n]))$.

Finally, note that $i_! \partial^+_n$ is exactly the unique natural transformation guaranteed by the universal property of the left Kan extension when applied to the composite $\partial \Delta_+[n] \xrightarrow{\partial^+_n} \Delta_+[n] \xrightarrow{\epsilon} \Delta[n]$ where $\epsilon$ is the unit of the relevant Kan extension. We also have $\partial \Delta_+[n] \xrightarrow{\xi} \partial \Delta[n] \xrightarrow{\partial_n} \Delta[n]$ where $\xi$ is again the unit of the relevant Kan extension.

By uniqueness we thus have that $\partial_n = i_! \partial^+_n$ and thus $i_! I_+ = I$ as required. \qed

**Lemma 4.5.4.** The counit of the adjunction $i_! \dashv i^*$ induces a weak equivalence on each of its components.

**Proof.** Let $X$ be any simplicial set and let $\epsilon_X: i_! i^* X \to X$ be the associated counit map. Firstly note that the unit of the above adjunction is always componentwise a weak equivalence as is shown in [154]. It is also well-known that the geometric realization of $i^* X$ is always homotopy equivalent to the normal geometric realization of $X$ (cf. [160], Proposition A.1.(iv).) This implies that the map $i^* \epsilon_X$ is a weak equivalence if and only if $\epsilon_X$ is. Now observe that the triangle identities for the adjunction give us $i^* \epsilon_X \circ \eta_{i^* X} = 1_{i^* X}$ and since both the identity on $i^* X$ and $\eta_{i^* X}$ are weak equivalences, by the thee-for-two
property of weak equivalences we get that \( i^* \epsilon_X \) is also a weak equivalence. Therefore, by the argument above, \( \epsilon_X \) is a weak equivalence, and we are done. \( \square \)

We are now ready to prove our main result.

**Proof of Theorem 1.** For necessity suppose \( S \) and \( T \) are homotopy equivalent. Since \( S \) and \( T \) are Kan complexes, there is a weak equivalence \( f: S \to T \) in \( \mathbf{sSets} \). By taking a fibrant replacement of \( f \) we can assume that \( f \) is also a fibration and therefore a trivial fibration. Trivial fibrations have the right lifting property with respect to cofibrations in any model structure but the cofibrations in \( \mathbf{sSets} \) are the monos (cf. [70], Proposition 3.2.2). Hence \( f \) has the right lifting property with respect to all monos and therefore in particular with respect to all boundary inclusions \( \partial_n \). Now let the following be a commutative diagram in \( \mathbf{ssSet} \):

\[
\begin{array}{ccc}
\partial \Delta_+[n] & \xrightarrow{g} & i^* S \\
\partial^+_n \downarrow & & \downarrow i^* f \\
\Delta_+[n] & \xrightarrow{h} & i^* T
\end{array}
\]

By applying \( i_! \), using Lemma 4.5.3 and post-composing with the counit \( \epsilon \) of the adjunction \( i_! \dashv i^* \) for each component we get the following commutative diagram in \( \mathbf{sSet} \):

\[
\begin{array}{ccc}
i_! \partial \Delta_+[n] = \partial \Delta & \xrightarrow{i!g} & i_! i^* S \xrightarrow{\epsilon_S} S \\
i_! \partial^+_n = \partial_n \downarrow & & \downarrow i_! i^* f \xrightarrow{\epsilon_T} f \\
i_! \Delta_+[n] = \Delta[n] & \xrightarrow{i!h} & i^* T \xrightarrow{\epsilon_T} T
\end{array}
\]

Since \( f \) is a trivial fibration by assumption, we get a lift \( \lambda: \Delta[n] \to S \) and therefore, by transferring along the adjunction, we get an arrow \( \hat{\lambda}: \Delta_+[n] \to i^* S \). The fact that \( \partial^+_n \circ \hat{\lambda} = g \) and \( i^* f \circ \hat{\lambda} = h \) follows immediately from the fact that \( \epsilon_S \circ i! g \) and \( \epsilon_T \circ i! h \) are the transposes of \( g \) and \( h \) respectively along the adjunction \( i_! \dashv i^* \). By Lemma 4.5.2
it thus follows that $i^*f$ is fiberwise surjective. Hence $i^*S$ and $i^*T$ are $\Delta^\text{op}_+\Delta^\text{op}$-equivalent as FOLDS-$\Delta^\text{op}_+$-structures.

For sufficiency, by Lemma 4.5.2, $i^*S \simeq_{\Delta^\text{op}_+} i^*T$ if and only if there exist $\eta: P \to i^*S$ and $\xi: P \to i^*T$ such that both have the right lifting property with respect to the class of maps $I_+$. It follows that $i\eta$ and $i\xi$ have the right lifting property with respect to the class of maps $iI_+$. By Lemma 4.5.3 we have that $iI_+ = I$ and since a map has the RLP with respect to $I$ iff it is a trivial fibration (cf. [70], Proposition 3.2.6 and Theorem 3.6.4) we get that both $i\eta$ and $i\xi$ are trivial fibrations and therefore weak equivalences.

We can now apply $i\iota$ and compose with the counits to get the following span from $S$ to $T$:

$$
\begin{array}{cccccc}
S & \xleftarrow{i^*S} & i\iota^*S & \xleftarrow{i^*\eta} & i\iota P & \xrightarrow{i^*\xi} & i\iota^*T & \xrightarrow{\epsilon_T} & T \\
\end{array}
$$

By Lemma 4.5.4 and three-for-two it is a span of weak equivalences, which implies the desired conclusion. $\square$

246
CHAPTER 5

Homotopy Model Theory: A Mathematical Logic for UF

As we have seen, the Univalent Foundations of Mathematics take their basic objects to be homotopy types. In UF mathematical structures are therefore encoded as structured homotopy types, similar to how in set-theoretic foundations they are encoded as structured sets. This basic picture allows us to envision a model theory in which the basic syntax no longer describes structured sets, but structured homotopy types. If “set-theoretic model theory” is understood as model theory in set-theoretic foundations then “homotopy type-theoretic model theory” would be the corresponding model theory in the Univalent Foundations. I have opted for the more sonorous “Homotopy Model Theory”. The present Chapter will develop the syntax and semantics of such a model theory, thus defining a usable mathematical logic for UF.

In [30] model theory is defined as “the branch of mathematical logic which deals with the relation between a formal language and its interpretations, or models.” Therefore, in order to do any kind of model theory we must at least have a formal language (syntax) and an interpretation of that formal language (semantics). This Chapter will develop these two components. They comprise what we will call “n-logic” (for $-1 \leq n < \infty$) in order to reflect the following fundamental (and revolutionary) idea of UF: an “$n$-level” theory is a piece of syntax whose models consist of structures defined on $n$-groupoids/homotopy $n$-types.

The syntax of $n$-logic will be based on a suitable extension of the syntax of Makkai’s FOLDS (First-Order Logic with Dependent Sorts) as it was developed in [110]. A fundamental insight of Makkai was that the syntax of “higher-level” theories can itself be
presented as a category. In particular, the signatures of FOLDS can be described as one-way categories where the arrows encode variable dependencies between the objects (understood as “sorts”). The key in defining \( n \)-logic is to “add equalities” to these FOLDS signatures. Indeed, it is helpful to think of \( n \)-logic as standing to FOLDS in the same relation that first-order logic with equality stands to first-order logic. To add equality to a first-order signature \( \Sigma \), one simply adds a binary relation with a certain fixed denotation. The analogous process for FOLDS is carried out in terms of a “globular completion” operation on categories which attaches (globular towers of) “equality sorts” to pre-existing sorts in \( \mathcal{L} \) in a manner compatible with their “height”. Roughly speaking, the signatures of \( n \)-logic are then the “globularly completed” FOLDS signatures of Makkai. This marks the first main contribution of this Chapter: a general definition of an “\( n \)-level” syntax.

The most natural semantics for \( n \)-logic is in UF. The basic idea of these “homotopy semantics” is the following: non-logical sorts of dimension \( m \leq n \) are interpreted as (dependent functions landing in) homotopy \( m \)-types, as the latter are formalized in (some) homotopy type theory (HoTT). The equality sorts that have been added to the syntax through globular completion are then interpreted as identity types. Although we will describe these semantics using the formal notation of dependent type theories (\( \Pi \), \( \Sigma \), Id-types etc.) we stress that no such fixed formal system is needed in order to make sense of the interpretation at an intuitive level. (The situation is entirely analogous to the set-theoretic semantics of first-order logic which can be understood at an intuitive level independent of any specific choice of formalism (e.g. ZFC, NBG) for the ambient set theory.) This marks the second main contribution of this Chapter: a semantics for \( n \)-logic that can be used to define structures in any HoTT.\(^1\)

\(^1\)For the purposes of this Chapter, HoTT will refer to the system outlined in [171], i.e. intensional MLTT with at least one univalent universe. But the semantics can easily be modified to accommodate alternative HoTTs, e.g. [21].
The semantics of $n$-logic can also be defined in a set-theoretic metatheory for suitable set-theoretic notions of $n$-groupoids/homotopy $n$-types. We will thus also define a set-theoretic semantics for 1-logic based on the more traditional notion of a groupoid understood as a category whose every morphism is invertible. These semantics interpret extensions of formulas in 1-logic as pseudonatural transformations from contexts to $\mathcal{L}$-structures. This marks the third main contribution of this Chapter: a groupoid semantics for a dependently-typed syntax entirely independent of the machinery of contextual/comprehension/type categories or $C$-systems [28, 71, 179] as well of the original groupoid interpretation of Hofmann and Streicher [68].

A natural question now is whether there is a proof system on the syntax of $n$-logic that can justify its homotopy semantics. In other words, even though the operation of globular completion adds “equality sorts” can we make their variables behave like paths in such a way that justifies their (fixed) denotation? We present such a proof system for $n$-logic. It is based on a standard sequent calculus for first-order logic to which we add three rules: two rules governing the existence and uniqueness of the “reflexivity” paths of equality sorts and a propositional version of the J-rule. Since the syntax of $n$-logic is purely relational (no closed terms) these assume an unfamiliar form, but they are in spirit very much related to the rules for identity types of MLTT [124]. This system is then shown to be sound with respect to both the above-sketched semantics, which marks the fourth main contribution of this Chapter.

Beyond the aims of this dissertation, we believe that there are two independent reasons to develop the model theory outlined above. Both relate to ongoing work in UF. Firstly, our framework provides a general definition of a signature for structures definable in the framework of UF. It could thus be used to generalize the Structure Identity Principle of [171]. Similar work has been carried out by Shulman, North and Ahrens [3] who
consider FOLDS in its capacity to provide a general notion of isomorphism for higher categories. Formalizations of category theory in the style of FOLDS has also been carried out by Ahrens under the UniMath project [181] and his formalization overlaps with some of the material of Section 5.7. Secondly, we intend $n$-logic as a tool for the study of homotopy type theories as mathematical objects themselves in such a way that can itself be formalized inside UF. In other words, $n$-logic can be used as a framework for doing metamathematics “natively” in UF, independent of any ambient set theory. This relates to ongoing work on the “Initiality Conjecture” in the series of papers by Voevodsky beginning with [179] (except we note that for its specific purposes this work takes place in ZF set theory). It also relates to work on “internal setoids” [144] which is focused on implementing tools to carry out the metatheory of type theories inside type theory.

The plan is as follows. In Section 5.1 we introduce the syntax of FOLDS. In Section 5.2 we introduce the operation of globular completion (Definition 5.2.1) and use it to define the syntax of $n$-logic for each (finite) $n$. In Section 5.3 we define the homotopy theoretic semantics of $n$-logic and in Section 5.4 we define a set-theoretic semantics for 1-logic. In Section 5.5 we define a proof system for $n$-logic and then in Section 5.6 prove soundness for both our semantics with respect to that proof system. Finally, in Section 5.7 we consider precategories as a FOLDS theory and then apply our framework to prove that univalent categories are “1-elementary” in the sense that they are axiomatizable by a 1-theory (Proposition 5.7.3).²

²This Chapter was presented at a Workshop on Categorical Logic in the University of Stockholm, Stockholm, December 2015, at the Univalent Foundations Seminar in the Institute for Advanced Study, Princeton, March 2016 and at a Workshop on Homotopy Type Theory and Categorical Logic (“HoTT16”) in the University of Leeds, Leeds, July 2016.
5.1. Preliminaries

We will assume familiarity with the basics of categorical logic as well as of intensional Martin-Löf Type Theory and its homotopy interpretation. For category-theoretic background [103] remains the standard reference; for the basics of dependent type theory we recommend [67, 124]; for the homotopy interpretation and an introduction to the Univalent Foundations see [77, 171] and references therein.

We will now present in more detail the basic syntax of FOLDS following closely Makkai’s original presentation in [110]. Our presentation will be given in “functorial” rather than “inductive” style. The apparent problem with the “functorial” style is that it invokes machinery much too strong to deserve the name “syntax”. However, all concepts introduced here can in fact be defined recursively with minimal assumptions since we impose appropriate finiteness conditions throughout.

**Definition 5.1.1 (FOLDS signature).** A (finite) FOLDS-signature is a category $\mathcal{L}$ with $|\text{ob}\mathcal{L}| < \aleph_0$ and $|\text{mor}\mathcal{L}| < \aleph_0$ together with a grading

$$d : \text{ob}\mathcal{L} \to \mathbb{Z}_{\geq -2}$$

such that for any $f : K \to K_f$ with $f \neq 1_K$ we have $d(K_f) > d(K)$ (where we write $K_f$ for the codomain of $f$). We call $d(K)$ the *dimension* of $K$. We define the *height* of $\mathcal{L}$ as

$$h(\mathcal{L}, d) = \sup_{K \in \text{ob}\mathcal{L}} d(K)$$

251
The dimension grading $d$ induces (canonically) a stratification of $\mathcal{L}$ into levels, defined (inductively) as follows:

$$l(K) = \begin{cases} 
0 & \text{if } K \text{ is the domain only of } 1_K \\
\sup_{f: K \to K_f} l(K_f) + 1 & \text{otherwise}
\end{cases}$$

We will call $l(K)$ the level of $K$.

**Remark 5.1.2.** It is not hard to see that Definition 5.1.1 is a generalization of Makkai’s definition of a FOLDS signature in [110]. In particular, the grading $d$ ensures that $\mathcal{L}$ is a finite, one-way and reverse well-founded category. Such categories are also known as inverse categories, except we restrict ourselves to finite ones. The dimension $d$ is extra structure not present in the original definition.

**Remark 5.1.3.** It is important to note that even though the level function is uniquely induced by $d$, it is possible to have objects that are of the same level, but of different dimension. For example consider $\mathcal{L}$ given by

$$
\begin{array}{c}
K_1 \\
\downarrow
\end{array} \quad \begin{array}{c}
\downarrow
\end{array} \quad \begin{array}{c}
K_2 \\
O
\end{array}
$$

with $d(K_1) = 2, d(K_2) = 3$ and $d(O) = 5$.

We will sometimes call an object $R$ of maximal level a relation and we call the rest of the objects of $\mathcal{L}$ sorts. (We will usually reserve the letter $R$ for the former and the letter $K$ for the latter.)
Example 5.1.4. Let \((\mathcal{L}_{\text{graph}}, d)\) denote the following FOLDS signature, with the numbers on the left representing the dimension of the corresponding sorts.

\[
\begin{array}{cc}
-1 & A \\
\uparrow & \downarrow \\
0 & O
\end{array}
\]

As the name suggests this would be the signature in which to talk about graphs, where \(A\) would encode the “edges” between previously declared “vertices” of sort \(O\). We will usually omit explicit mention of \(d\), and write simply \(\mathcal{L}_{\text{graph}}\) for the above signature.

Example 5.1.5. Similarly, omitting \(d\), we let \(\mathcal{L}_{\text{rg}}\) denote the following signature

\[
\begin{array}{cc}
-1 & I \\
\uparrow & \downarrow i \\
0 & A \\
\downarrow c & \downarrow d \\
1 & O
\end{array}
\]

subject to the relation \(d i = c i\). Intuitively, this corresponds to the signature for reflexive graphs, where \(I\) is a unary predicate that can only be “asked” of an “arrow” in \(A\) that we already know is a loop. \(\mathcal{L}_{\text{rg}}\) will serve as a fundamental example since it represents the simplest case sufficient to illustrate most of the concepts we are going to introduce.

Another notion that is crucial for us is that of a (finite) extension of one FOLDS-signature by another. Recall that a cosieve in a category \(\mathcal{C}\) is a class of maps closed under postcomposition and if \(\mathcal{C}'\) is a subcategory of \(\mathcal{C}\) we say that \(\mathcal{C}'\) is a cosieve in \(\mathcal{C}\) if \(\operatorname{mor}\mathcal{C}'\) is a cosieve as a class of maps in \(\mathcal{C}\).
Definition 5.1.6 (Extension of a signature). Let \((\mathcal{L}, d)\) be a FOLDS signature. Then a FOLDS signature \((\mathcal{L}', d')\) is an extension of \((\mathcal{L}, d)\) iff \(\mathcal{L}\) is a cosieve in \(\mathcal{L}'\) and \(d'|_{\text{ob}\mathcal{L}} = d\).

We write \((\mathcal{L}', d') > (\mathcal{L}, d)\) to indicate that \((\mathcal{L}', d')\) is an extension of \((\mathcal{L}, d)\).

For the rest of this section we assume we are given a FOLDS signature \((\mathcal{L}, d)\). We will now define a language out of it. By a language here we mean the collection of all concepts standardly associated with a syntax: variables, formulas, sequents and rules for determining the well-formedness of each. (At this point the dimension \(d\) plays no essential role other than inducing the level function \(l\).)

Definition 5.1.7 (Variables and Contexts). Variables are given by a functor

\[ V : \mathcal{L} \to \text{Set} \]

satisfying the following conditions:

1. \(V(K) \cap V(K') = \emptyset\) for \(K \neq K' \in \text{ob}\mathcal{L}\)
2. \(|V(K)| = \aleph_0\) for all \(K \in \text{ob}\mathcal{L}\)
3. For every finite (i.e. finite sets in all its values) subfunctor \(\Gamma \subset V\) we have

\[ |\{y \mid \text{dep}(y) \subset |\Gamma|\}| = \aleph_0 \]

where

\[ \text{dep}(y) = \text{def} \{V(f)(x)\mid \text{dom}(f) = K\} \]

is the set of dependent variables of \(x\) and

\[ |\Gamma| = \bigcup_{K \in \mathcal{L}} \Gamma(K) \]

254
We write \( x: K \) as an abbreviation for \( x \in V(K) \). A context is a finite subfunctor of \( V \) and given two such contexts \( \Gamma, \Delta \) a context morphism is a natural transformation between them. For any two contexts \( \Gamma \) and \( \Delta \) we write \( \Gamma \cup \Delta \) for their union as subobjects of \( V \), i.e. the functor that takes \( K \mapsto \Gamma(K) \cup \Delta(K) \). If \( \Gamma \) is a context and \( x: K \) a variable such that \( \text{dep}(x) \subset \Gamma \) then we write \( \Gamma, x: K \) for the context that is the same as \( \Gamma \) everywhere except \( \Gamma, x: K(K) = \Gamma(K) \cup \{x\} \). Like with contexts we write

\[
|V| = \bigcup_{K \in \mathcal{L}} V(K)
\]

For a given context \( \Gamma \) we write

\[
\Gamma^\uparrow = \{ y \in |V| | y \notin \text{dep}(x) \forall x \in |\Gamma| \}
\]

When \( x \in \Gamma^\uparrow \) and \( x \notin \Gamma \) we say that \( x \) is fresh for \( \Gamma \).

**Remark 5.1.8.** Condition (3) in Definition 5.1.7 is there to ensure that there are always enough variables available to us given any choice of other variables that they may depend on. Since the notions we use to express condition (3) are also introduced in terms of \( V \) there is clearly a circularity here. But it is one that can easily be circumvented with a mutually inductive definition. How to actually implement such a definition is of course a non-trivial technical problem.

It is easy to check that the notions in Definition 5.1.7 encode the same amount of information as the corresponding notions in MLTT. We will usually write out contexts in the MLTT style. We will also often abbreviate contexts by writing out simply the names of variables in them.
Example 5.1.9. Let $f \in V(A)$ and $x, y \in V(O)$ and $V(d)(f) = x$, $V(c)(f) = y$ and define the context $\Gamma: \mathcal{L}_{\text{rg}} \to \textbf{Set}$ by

$$\Gamma(I) = \varnothing, \Gamma(A) = \{f\}, \Gamma(O) = \{x, y\}$$

and the obvious action of the maps on $f$. Then we can think of $\Gamma$ as the context

$$\{x: O, y: O, f: A(x, y)\}$$

which we will abbreviate to $\{x, y, f\}$.

Example 5.1.10. An important example of a context morphism is given by the “contraction” morphism

$$\{x: K, y: K, p: x =_K y\} \to \{x: K, q: x =_K x\}$$

that sends both $x$ and $y$ to $x$.

We can now define the set of $\mathcal{L}$-formulas and their associated contexts of free variables by simultaneous induction.

Definition 5.1.11 (Formulas and Sequents). $\top$ and $\bot$ are atomic formulas and $\text{FV}(\top) = \text{FV}(\bot) = \varnothing$. If $\phi$, $\psi$ are formulas then so are $\phi \land \psi$, $\phi \lor \psi$ and $\phi \to \psi$ with

$$\text{FV}(\phi \land \psi) = \text{FV}(\phi \lor \psi) = \text{FV}(\phi \to \psi) = \text{FV}(\phi) \cup \text{FV}(\psi)$$

Negation $\neg \phi$ can be defined as $\phi \to \bot$. For the quantifiers, assume that $\phi$ is a formula and that $x$ is a variable of sort $K$ in $\text{FV}(\phi)$. Then $\forall x: K. \phi$ and $\exists x: K. \phi$ are formulas.
provided that \( x \in \text{FV}(\phi) \uparrow \). In that case we have

\[
\text{FV}(\forall x: K.\phi) = \text{FV}(\exists x: K.\phi) = \text{FV}(\phi) \cup \text{dep}(x) \setminus \{x\}
\]

We define a \textit{sequent} as a syntactic entity of the form

\[ \Gamma \vdash \phi \vdash \psi \]

where \( \phi, \psi \) are \( \mathcal{L} \)-formulas and \( \Gamma \supset \text{var}(\phi) \cup \text{var}(\psi) \). (We follow [71] in using the notation \( \Gamma \vdash \phi \vdash \psi \) in order to avoid overloading subscripts.)

\textbf{Remark 5.1.12.} The justification for not having top-level sorts correspond to atomic formulas in the form of relation symbols is that it simplifies proofs about the whole syntax by allowing us to consider fewer cases. From now on we will allow ourselves to switch to the usual “sugared” form whenever convenient. In \( \mathcal{L}_{rg} \), for example, \( I(f, x) \) will be syntactic sugar for \( \exists \tau: I(f, x).T \).

\textbf{Remark 5.1.13.} For reasons that will become clear in Section 5.5, we can without loss of generality restrict ourselves to “singular” sequents, i.e. those that consist of single formulas on either side of the turnstile. For added generality, most of our proofs below will be carried out in terms of sequents rather than individual formulas, although this changes nothing of essence in our arguments.

It remains to define substitution. Let \( \Gamma, \Delta \) be (well-formed) contexts and \( s: \Gamma \Rightarrow \Delta \) a context morphism. We need to describe how such a context morphism acts on formulas. Let \( \phi \) be a formula in context \( \Gamma \). Then we define \( s(\phi) \), the substitution of \( \phi \) along \( s \), as follows:

- If \( \phi \equiv T, \bot \) then \( s(\phi) \equiv T, \bot \)
• If $\phi \equiv \psi \land \chi, \psi \lor \chi, \psi \rightarrow \chi$ then $s(\phi) \equiv s(\psi) \land s(\chi), s(\psi) \lor s(\chi), s(\psi) \rightarrow s(\chi)$

• If $\phi \equiv \exists x: K.\psi, \forall x: K.\psi$ then $s(\phi) \equiv \exists y: K.s(\psi), \forall y: K.s(\psi)$ where $y$ is fresh for $\Gamma \cup \Delta \cup \{x\}$ such that $\text{dep}(y) = s(\text{dep}(x))$.

Clearly, in the last clause, there are many distinct choices of $y$. As Makkai also notes, this makes $s(\phi)$ not a well-defined operation. There are many ways of rectifying this, e.g. by imposing a well-ordering on variables or by defining the action of $s$ on $\alpha$-equivalence classes of formulas rather than formulas themselves. Since we are primarily concerned with describing $n$-logic at a high level of generality, we will simply assume that we have chosen one such way of making substitution well-defined on formulas. (Needless to say, different choices of achieving this may be available to us depending on whether we are using set theory or HoTT as our metatheory.)

5.2. Syntax of $n$-logic

We will now introduce the syntax of $n$-logic as an extension of the syntax of FOLDS described in Section 5.1. The key construction here is an operation that “adds equalities” to FOLDS signatures. This operation is suggestively called globular completion. After this is done, we are allowed to “extend” globularly completed signatures by new sorts that may depend on the equality sorts we just added. The signatures of $n$-logic are then obtained as arbitrarily large finite iterations of this process of first globularly completing and then extending. Given these signatures, the syntax is then defined as in Section 5.1. The syntax of $n$-logic is thus a formalization of the idea that we “add equalities” (globular completion) and then by extending add predicates and relations that talk about these equalities.

**Definition 5.2.1 (Globular Completion).** Let $(L, d)$ be a FOLDS signature of height $n$. The *globular completion* $(L^\equiv, d^\equiv)$ of $(L, d)$ is given by the following data:
(1) $\mathcal{L}^=\text{ contains all of } \mathcal{L}$

(2) For each $K \in \mathcal{L}$ with $d(K) > -1$, $\mathcal{L}^=\text{ contains kinds } =^1_K, ..., =^{d(K)+1}_K \text{ and arrows}$

\[ s^K_i, t^K_i : =^i_K \to =^{i-1}_K \]

that satisfy the globular identities

\[ s^K_{i-1} \circ s^K_i = t^K_{i-1} \circ t^K_i \]

\[ t^K_{i-1} \circ s^K_i = t^K_{i-1} \circ t^K_i \]

(3) For any $K \in \mathcal{L}$ with $0 \leq d(K) < n$ and any $f : K \to K'$ we add the relation

\[ f \circ s_1 = f \circ t_1. \]

(4) For each $=^j_K$ with $j \leq n$ we add a sort $r^K_j$ and an arrow

\[ \rho^K_j : t^K_j \to =^j_K \]

such that

\[ s^K_j \circ \rho^K_j = t^K_j \circ \rho^K_j \]

(5) We define $d^=\text{ as follows:}$

- $d^= (K) = d(K) \text{ for all } K \in \text{ob}\mathcal{L}$
- $d^= (=^i_K) = d(K) - i$
- $d^= (r^K_j) = d(K) - (i + 1)$

$(\mathcal{L}^=, d^=)$ is thus a FOLDS signature of height $n$ (with extra structure in the form of specified sorts). We will call these new sorts \textit{logical sorts}.
Remark 5.2.2. Clearly, \( s \) and \( t \) are to be understood as “source” and “target” maps for equality “paths”. On the other hand, \( r \) is to be understood as the predicate picking out the “reflexivity” proof of equality and the identities are there to ensure that \( r \) only applies to those equality sorts declared with identical variables. We will usually suppress explicit mention of \( K \) in the \( s, t \) and \( r \).

Remark 5.2.3. The reflexivity sorts of dimension \(-2\) play essentially no role in the syntax and we will usually omit writing them out in our examples. They are there to ensure the reflexivity of equality sorts of dimension \(-1\). The reason we add them is that they allow a more uniform presentation of the proof system in Section 5.5 in that all the properties of equality can be expressed by single set of rules.

Remark 5.2.4. We will also stipulate that the globular completion operation is idempotent. Once it has already been applied to \((\mathcal{L}, d)\), reapplying it does nothing.

Example 5.2.5. Consider \( \mathcal{L}_{rg} \). Then \( \mathcal{L}_{rg}^- \) is the following signature
subject to the following extra relations:

\[
\begin{align*}
  s_1^O \circ s_2^O &= s_1^O \circ t_2^O, t_1^O \circ s_2^O = t_1^O \circ t_2^O \\
  t_1^O \circ \rho_1^O &= s_1^O \circ \rho_1^O, c \circ t_1^A c &= c \circ s_1^A c \\
  t_2^O \circ \rho_2^O &= s_2^O \circ \rho_2^O, t_1^A \circ \rho_1^A &= s_1^A \circ \rho_1^A \\
  d \circ t_1^A c &= d \circ s_1^A c
\end{align*}
\]

Crucially, extensions of globularly completed signatures allow us to express properties and impose structure on the identity sorts themselves. For example, the following is an extension of \( \mathcal{L}_{rg} \)

\[
\begin{align*}
  \text{Diagram here}
\end{align*}
\]

We will put this idea to use in axiomatizing univalent categories (Proposition 5.7.3).

**Remark 5.2.6.** Note that isomorphic categories with different dimension functions will generally give rise to very different globular completions, even if they are of the same height. For example, the signature \(( \mathcal{L}_{graph}, d \) with \( d(O) = 2 \) and \( d(A) = 0 \)) will give a different globular completion from \(( \mathcal{L}_{graph}, d' \) with \( d'(O) = 2 \) and \( d'(A) = 1 \). Syntactically, one can think of the dimension of each sort as specifying how “tall” its globular completion will be. Semantically, as we shall see, the dimension determines the given sort’s \( h \)-level.
Let us write $\Lambda_n^0$ for the set of FOLDS signatures of height $n$. We define

$$\Lambda_n^1 = \{(\mathcal{L}^-, d^-)|(\mathcal{L}, d) \in \Lambda_n^0\}$$

and

$$\bar{\Lambda}_n^1 = \{(\mathcal{L}', d')|(\mathcal{L}', d') > (\mathcal{L}, d) \in \Lambda_n^1\}$$

where $>$ is the relation of extension in Definition 5.1.6. We can now write

$$\Lambda_{n+1}^i = \{(\mathcal{L}^-, d^-)|(\mathcal{L}, d) \in \bar{\Lambda}_n^i\}$$

for arbitrary $0 \leq i \in \mathbb{N}$ and then define

$$\Lambda_n = \bigcup_{i \in \mathbb{N}} \Lambda_n^i$$

$\Lambda_n$ is the set of signatures for $n$-logic, or $n$-signatures. These signatures are FOLDS signatures in the sense of Section 5.1, except they also contain extra structure in the form of the logical sorts $=^i_K, r^i_K$. We then define the syntax of $n$-logic simply to be the FOLDS syntax for signatures in $\Lambda_n$. Thus, given an $n$-signature $\mathcal{L}$, the $\Lambda_n$-formulas (resp. sentences, sequents etc.) for $\mathcal{L}$ are simply the FOLDS $\mathcal{L}$-formulas (resp. sentences, sequents etc.) as described in Section 5.1.

**Remark 5.2.7.** It is straightforward to see that if $(\mathcal{L}, d) \in \Lambda_n$ then $(\mathcal{L}, d) \in \Lambda_n^i$ for some $i \leq n + 1$ and so $\Lambda_n$ can be obtained only after a finite number of steps. (We could have avoided the redundancy, for example, by defining extensions as only those signatures that add sorts dependent on previously-introduced equality sorts.)

**Remark 5.2.8.** $\Lambda_n$ can straightforwardly be described as the (category of) algebras of a monad on the category of FOLDS signatures (with possibly specified logical sorts) and
morphisms the dimension-preserving extensions that preserve these logical sorts on the nose. In future work the “monadic packaging” of the syntax will likely prove necessary if more general results are to be proved for $n$-logic involving higher $n$ (e.g. a general completeness theorem).

5.3. Homotopy Semantics of $n$-logic

We will write out the semantics for $n$-logic in the syntax of (intensional) MLTT with homotopy type theory as our metatheory. (HoTT is here understood as intensional MLTT with at least one univalent universe; indeed we can take 1UHoTT as described in the Appendix of Chapter 3 as our metatheory as long as we make sure that we have enough expressive power to truncate types in the manner our semantics requires.) More generally, our semantics could be defined in any dependent type theory with a good notion of $h$-levels, or indeed in any categorical model of such a type theory (e.g. contextual categories [28] or C-systems [179]).

NOTATION. We will write $\mathcal{U}$ for a univalent universe of types and $||A||$ for the propositional truncation of a type $A$ in $\mathcal{U}$. For any type $A$ we will write $\text{Id}_A^i$ for the $i$th-iterated identity type. Thus, for example, $\text{Id}_A^3(\alpha, \beta)(p, q)(x, y)$ stands for

$$\text{Id}_{\text{Id}_{A^3}(x, y)(p, q)}(\alpha, \beta)$$

Otherwise we will follow the notation of [171] closely.

Fix $n \geq -1$ and $(\mathcal{L}, d) \in \Lambda_n$. We write $K_j^j$ for each object in $\mathcal{L}$ where $j = l(K_j^j)$. Similarly, we write $f_{ijkl}$ for each morphism $K_j^j \to K_k^l$ in $\mathcal{L}$. For each $K_j^j$ in $\mathcal{L}$ we write $a_{ijkl} \in \mathbb{N}$ for the cardinality of the set of all non-identity morphisms from $K_j^j \to K_k^l$, i.e. $a_{ijkl} = |\mathcal{L}(K_j^j, K_k^l)|$. We write $a_{ij}$ for the cardinality of the set of all non-identity morphisms out of $K_j^j$ in $\mathcal{L}$ and $A_{ij}$ for that set of morphisms, i.e. $a_{ij} = |A_{ij}|$. It will
also be convenient to write $D_{ij}$ for the set of lower indices of the codomains of each morphism in $A_{ij}$, i.e. $D_{ij} = \{ k | f_{ijkl} \in A_{ij} \}$. We will also assume there is an ordering on $A_{ij}$ such that we have an induced order-preserving indexing $p^{ij}: \{ 1, ..., a_{ij} \} \to A_{ij}$.

Thus when we write $p^{ij}_k$ this means the $k$-th morphism out of $K^j_i$ for the given ordering.

Similarly, we will assume that there is an indexing of the codomains of these morphisms $d^{ij}: \{ 1, ..., a_{ij} \} \to D_{ij}$ such that $d^{ij}_k$ is the lower index of the codomain of $p^{ij}_k$.

The basic idea of an $\mathcal{L}$-structure is as follows. First we give the denotation of the non-logical sorts by induction on their level: “bottom-level” sorts $K$ are $n$-types and sorts of level $m \geq n$ are dependent functions into $d(K)$-types which rely on the definitions of the non-logical sorts of lower level. Logical sorts are then defined as the identity sorts and reflexivity predicates on sorts that have already received denotation. We then assign denotations to sorts that may depend on logical sorts. We repeat the process until everything has been assigned a denotation. The following (long) definition spells out this process.

**Definition 5.3.1 (Homotopy $\mathcal{L}$-structure).** An $\mathcal{L}$-structure $\mathcal{M}$ (or $\mathcal{L}$-$n$-structure if we want to make $n$ explicit) is obtained by the following process. We begin with non-logical sorts none of which depend on logical sorts (i.e. those that belong to the initial FOLDS signature, before it is globularly completed). We assign them denotations by induction on their level as follows:

1=0 For each $K^0_i$ we pick an $n$-type $\mathcal{M}(K^0_i): n\text{-type}_d$

1=1 For each $K^1_i$ we pick a term

\[ \mathcal{M}(K^1_i): \mathcal{M}(K^0_{p^1_1}) \to \mathcal{M}(K^0_{p^2_1}) \to \ldots \to \mathcal{M}(K^0_{p^a_{i1}}) \to d(K^1_i)\text{-type}_d \]
For each $K_i^2$ we pick a term

$$\mathcal{M}(K_i^2): \prod_x \mathcal{M}(K_{d_1}^1)(x_{d_1}^{p_1}, \ldots, x_{d_1}^{p_{i-1}}) \to \mathcal{M}(K_{d_2}^1)(x_{d_2}^{p_1}, \ldots, x_{d_2}^{p_{i-1}})$$

$$\to \ldots \to \mathcal{M}(K_{d_{n-1}}^1) \to d(K_i^2)\text{-type}_U$$

Beyond $l = 2$ it becomes very difficult to write down the relevant types while keeping all variable dependencies explicit. Suppressing variable dependencies we can thus write the data for a sort of level $n$ (equivalently, of dimension $-1$) as follows:

$$\mathcal{M}(K_i^n): \prod_{\bar{x}_{d_1}^{p_1}} \mathcal{M}(K_{d_1}^{n-1}) \to \ldots \to \mathcal{M}(K_{d_{n-1}}^{n-1}) \to (-1)\text{-type}_U$$

On the other hand, for logical sorts we do the following, writing $K^M$ for the denotation of $K$ in $\mathcal{M}$ and $\Gamma_K^M$ for its canonical context. Then for any non-logical $K \in \text{ob}\mathcal{L}$ we let $(\equiv_i^K)^M$ be the $i$th-iterated identity type “over” $K^M$:

$$(\equiv_i^K)^M \equiv \lambda \bar{x}: \Gamma_K^M. \lambda y, z: K^M(\bar{x}).$$

$$\lambda p_1, q_1: \text{Id}^M(y, z) \ldots \lambda p_i, q_i: \text{Id}^{M(\bar{x})}(p_{i-1}, q_{i-1}) \ldots (p_1, q_1)$$

$$: \prod_{\bar{x}: \Gamma_K^M} (d(K) - i)\text{-type}_U$$

For reflexivity sorts we define

$$r_i^i \equiv \lambda \bar{x}: \Gamma_K^M. \lambda y, z: K^M(\bar{x}) \ldots \lambda q: \text{Id}^{M(\bar{x})}(y, z) \ldots (p_{i-2}, q_{i-2}).$$

$$\lambda p: \text{Id}^{M(\bar{x})}(y, z) \ldots (q, q). \text{Id}(p, \text{refl}_q): \prod_{\bar{x}: \Gamma_K^M} (d(K) - (i + 1))\text{-type}_U$$

$$p_j, q_j: \text{Id}^{M(\bar{x})}(y, z) \ldots (p_{j-1}, q_{j-1})$$

$$1 \leq j \leq i - 2$$

$$q: \text{Id}^{M(\bar{x})}(y, z) \ldots (p_{i-2}, q_{i-2})$$

$$p: \text{Id}^{M(\bar{x})}(y, z) \ldots (q, q)$$

265
Thus all “first-generation” sorts and their identity sorts and reflexivity relations have been assigned denotation. We now repeat the same process for later generations until everything has been assigned a denotation. This completes the definition of an $\mathcal{L}$-structure $\mathcal{M}$.

**Remark 5.3.2.** Clearly, the definition of an $\mathcal{L}$-structure is not functorial in the sense that the relations in $\mathcal{L}$ encoding variable dependencies are “written in” as judgmental equalities by hand, rather than us first coding up $\mathcal{L}$ as a category in HoTT and defining an $\mathcal{L}$-structure as a functor out of it. (Another way of saying this is that our definition of an $\mathcal{L}$-structure is given in the “Reedy way”.) This is a well-known limitation in intensional MLTT that hinders the uniformity of our definition. We console ourselves in the fact that given a finite signature $\mathcal{L}$ it will always in principle be possible to carry out the process described in Definition 5.3.1. Doing so efficiently (even algorithmically) is another matter altogether.

**Remark 5.3.3.** The fact that $(=^i_K)^\mathcal{M}$ is indeed a term of the above-specified type depends on us having asserted that each $K^\mathcal{M}(\vec{x})$ will be a $d(K)$-type (i.e. that we have a proof of that fact in hand) and similarly for the denotation of reflexivity sorts.

**Remark 5.3.4.** An $\mathcal{L}$-$n$-structure $\mathcal{M}$ may depend on some (non-empty) ambient context (i.e. be an “interpretation with parameters”). In that case we will denote such an ambient context by $\Delta^\mathcal{M}$. For example, in an $\mathcal{L}$-$n$-structure-with-parameters $\mathcal{M}$, the type declarations of ground sorts $K$ are to be understood as claiming that the following judgment is derivable

$$\Delta^\mathcal{M} \vdash \mathcal{M}(K) : \text{n-type}_U$$

where, as before, we abuse notation in identifying the pair $K$ with its first projection.
Example 5.3.5. Let \( \mathcal{L} = \mathbf{1} \), where \( \mathbf{1} \) is the category with one object and one identity arrow. Then an \( \mathbf{1} \-(-1) \)-structure is simply a mere proposition \( P : \text{Prop}_\mathcal{U} \). This allows us to say that \((-1)\)-logic has the same expressive power as propositional logic. Similarly, a \( \mathbf{1} \-n \)-structure is simply an \( n \)-type. Thus the empty \( \mathbf{1} \-n \)-theory is simply the theory of \( n \)-groupoids (understood as \( n \)-types in HoTT). This provides a generalization of the fact that the empty theory over a signature with a single sort in first-order logic is simply the theory of sets, also known as the “pure theory of identity”.

Example 5.3.6. An \( \mathcal{L}_{\text{graph}} \)-structure consists of an \( h \)-set \( O : \mathcal{U} \) and a mere relation \( A : O \to O \to \text{Prop}_\mathcal{U} \). More generally, the study of \( 0 \)-signatures coincides with traditional set-based model theory (with the restriction that we are only considering relation symbols). \( 0 \)-logic can thus be thought of as the “classical limit” of \( n \)-logic. Consider \( \mathcal{L}^=_{\text{graph}} \) as an illustration:

\[
\begin{array}{ccc}
-1 & A & =_O \\
\downarrow & \downarrow & \downarrow \\
0 & O & \\
\end{array}
\]

\( \mathcal{L}_{\text{graph}} \)-0-formulas are then exactly (when suitably translated) the formulas of first-order logic with equality for a single-sorted signature \( \Sigma \) with a single binary predicate \( A \). Semantically, an \( \mathcal{L}_{\text{graph}} \)-0-structure \( \mathcal{M} \) consists of a \( 0 \)-type \( O^\mathcal{M} : \text{Set}_\mathcal{U} \) and a dependent type \( A^\mathcal{M} : O^\mathcal{M} \to O^\mathcal{M} \to \text{Prop}_\mathcal{U} \). This is all entirely analogous to \( \Sigma \)-structures in traditional set-theoretic semantics. We may say that \( 0 \)-logic has the same expressive power as first-order logic with equality and its semantics as defined here coincide with the usual set-theoretic semantics.
Example 5.3.7. An $\mathcal{L}_{\text{rg}}$-1-structure $\mathcal{M}$ consists of the following data:

$$I^\mathcal{M}: \Pi_{x:O^\mathcal{M}} A(x,x) \to \text{Prop}_\mathcal{U}$$

$$A^\mathcal{M}: O^\mathcal{M} \to O^\mathcal{M} \to \text{Set}_\mathcal{U}$$

$$O^\mathcal{M}: \text{Gpd}_\mathcal{U}$$

Given a context $\Gamma$ (in $\mathcal{L}$) written in the MLTT style as

$$\Gamma = \{x_1: K_1, x_2: K_2(x_1), \ldots, x_n: K_n(x_1, \ldots, x_{n-1})\}$$

we obtain a well-formed context in MLTT as follows

$$\Gamma^\mathcal{M} = \{x_1: K_1^\mathcal{M}, \ldots, x_n: K_n^\mathcal{M}(x_1, \ldots, x_{n-1})\}$$

where we keep the variable names the same for convenience. (When $\Gamma$ is empty we obviously take $\Gamma^\mathcal{M}$ also to be the empty context.) A context morphism (in MLTT) $\vec{a}: \Delta^\mathcal{M} \to \Gamma^\mathcal{M}$ is then called a realization of $\Gamma$ in $\mathcal{M}$.

With this in mind, we now define the interpretation of $\mathcal{L}$-formulas in an $\mathcal{L}$-$n$-structure $\mathcal{M}$. This proceeds much as the PAT-style described in Definition 3.3.1: universal quantifiers become $\Pi$-types and existential quantifiers become $\Sigma$-types and conjunction, disjunction, implication and negation are translated in the usual manner, namely as the type formers $\times$, $+$, $\to$ and $(-) \to \mathbf{0}$ respectively.
**Definition 5.3.8 (Interpretation of Formulas).** Let $\phi$ and $\psi$ be $\mathcal{L}$-formulas. We interpret them as types in HoTT as follows:

\[
\begin{align*}
\top^\mathcal{M} &= \text{def} \ 1 \\
\bot^\mathcal{M} &= \text{def} \ 0 \\
(\phi \land \psi)^\mathcal{M} &= \text{def} \ \phi^\mathcal{M} \times \psi^\mathcal{M} \\
(\phi \lor \psi)^\mathcal{M} &= \text{def} \ ||\phi^\mathcal{M} + \psi^\mathcal{M}|| \\
(\phi \rightarrow \psi)^\mathcal{M} &= \text{def} \ \phi^\mathcal{M} \rightarrow \psi^\mathcal{M} \\
(\neg \phi)^\mathcal{M} &= \text{def} \ \phi^\mathcal{M} \rightarrow 0 \\
(\exists x : K. \phi(x))^\mathcal{M} &= \text{def} \ ||\sum_{x : \mathcal{M}} \phi^\mathcal{M}|| \\
(\forall x : K. \phi(x))^\mathcal{M} &= \text{def} \ \Pi_{x : \mathcal{M}} \phi^\mathcal{M}
\end{align*}
\]

**Remark 5.3.9.** There is a certain ambiguity in Definition 5.3.8 since what is on the right hand side is meant to be interpreted as a type, yet it is possible that there are free variables floating around. Thus, the above definitions are better understood as saying that formulas are interpreted as judgements that the right-hand side is a type in some context that contains its free variables.

**Example 5.3.10.** In $\mathcal{L}_{\text{rg}}$ if we are given the formula $\phi \equiv \exists \tau : I(x,f).\top$ then its interpretation in some $\mathcal{L}_{\text{rg}}$-structure $\mathcal{M}$ will be given by $||\sum_{\tau : I^\mathcal{M}(x,f)} 1||$ which is of course equivalent to $||I^\mathcal{M}(x,f)||$. But this is not strictly speaking a type since $x$ and $f$ remain variables. Instead, we should take the interpretation of $\phi$ in $\mathcal{M}$ to consist of the following
judgement

$$\Delta^{M}, x: O^{M}, f: A^{M}(x, x) \vdash ||I^{M}(x, f)||: U$$

where we are abusing notation in using $I^{M}(x, f)$ for what really is its first projection since strictly speaking $I^{M}$ was defined as a dependent function into $\text{Prop}_U$. Since both $0$ and $1$ are types in the empty context, by the weakening rule for contexts we will get an analogous definition for any formula $\phi$ in any context, possibly larger than its context of free variables.

Remark 5.3.11. As the above example illustrates, all non-trivial propositions are to be constructed using the existential quantifier. As is reasonable, we will refrain from interpreting a formula $\exists x: A. \top$ as $|| \sum_{x: A^{M}} 1 ||$ and write it instead as $||A^{M}||$. The latter two types are of course equivalent.

We are now ready to define a notion of satisfaction for formulas and sequents.

Definition 5.3.12 (Satisfaction). Let $\phi$ be any $\mathcal{L}$-formula, $\mathcal{M}$ any $\mathcal{L}$-structure, $\vec{a}$ any realization of a context $\Gamma \supset \text{FV}(\phi)$ in $\mathcal{M}$. We define satisfaction of $\phi$ by $\vec{a}$ in $\mathcal{M}$ as follows:

$$\mathcal{M} \models [\phi[\vec{a}/\Gamma]] \text{ iff } \phi^{M}[\vec{a}/\Gamma] \text{ is inhabited}$$

The case where $\phi$ has no free variables is a special case of the above definition, in which case we write $\mathcal{M} \models \phi$ and say that $\mathcal{M}$ is a model of $\phi$. Satisfaction for sequents can be defined similarly and is suggested by their very notation: For sequents we define satisfaction as follows:

$$\mathcal{M} \models \Gamma \models \phi \iff \Delta^{M}, \Gamma^{M}, x: \phi^{M} \vdash y: \psi^{M} \text{ is derivable}$$
Remark 5.3.13. The right-hand side of the first biconditional in Definition 5.3.12 is equivalent to requiring that a judgement

\[ \Delta^M \vdash \pi: \phi^M[\bar{a}/\Gamma] \]

is derivable in HoTT, where \( \pi \) is a (metavariable for a) term of the relevant type.

Remark 5.3.14. Since we will always assume the presence of \( \Pi \)-types in our metatheory, the right-hand side of the second biconditional in Definition 5.3.12 is equivalent to the statement that the type

\[ \Pi_{\bar{x}:\Gamma^M} \phi^M \to \psi^M \]

is inhabited (in context \( \Delta^M \)). We will most often interpret sequents in this latter form, since it is the more convenient (albeit less general) of the two.

Example 5.3.15. Consider the following \( \mathcal{L}_{rg} \)-sentence

\[ \phi \equiv \forall x: O \exists f: A(x,x).I(f) \]

and let \( \mathcal{M} \) be the \( \mathcal{L}_{rg} \)-1-structure given by the following data

\[ \langle \text{Set}_U, \lambda x.\lambda y. x \to y, \lambda x.\lambda f. \text{Id}_{x \to x}(f,1_x) \rangle \]

with ambient context \( \emptyset \). Then \( \mathcal{M} \) is a model of \( \phi \). To see this, observe that

\[ \mathcal{M} \models \forall x: O \exists f: A(x,x).I(f) \]

since

\[ \emptyset \vdash \lambda x.(1_x, \text{refl}_{1_x}): \prod_{x: \text{Set}_U} \Sigma f: x \to x \text{Id}_{x \to x}(f,1_x) \]
is derivable in UF and since

\[ \forall x : O \exists f : A(x, x).I(f) \equiv \prod_{x : \text{Set}_{/x}} \Sigma_{f : x \to x} \text{Id}_{x \to x}(f, 1_x) || \]

Finally, note that the way we have set up our semantics all the types that interpret sentences and substitution instances of formulas will be \(h\)-props (“mere propositions”). This is because every atomic formula is a mere proposition and all logical constructors preserve \(h\)-props except + and \(\Sigma\). But for the latter we have taken their propositional truncation, as is also done in [171]. We can record this fact as a proposition, since the soundness of our proof system \(D\) introduced below crucially depends on it.

**Proposition 5.3.16.** For any \(L\)-formula \(\phi\), \(L\)-n-structure \(M\), context \(\Gamma \supset \text{FV}(\phi)\) and realization \(\bar{a}\) of \(\Gamma\) in \(M\) we have that \(\phi^M[\bar{a}/\Gamma]\) is a mere proposition.

### 5.4. Set-Theoretic semantics for 1-logic

The semantics we have presented in Section 5.3 use HoTT as a metatheory and thus take place within the Univalent Foundations. In this section we give a semantics for \(n\)-logic in set theory. The easiest way of doing this is would be simply to interpret the syntax of \(n\)-logic in a category with the appropriate structure to interpret the required type constructors (e.g. \(C\)-systems or contextual categories) or even directly into the category of Kan complexes in simplicial sets, following the “canonical” model of homotopy type theory constructed in [77]. However, such a semantics does not add anything essentially new to the homotopy semantics presented in Section 5.3. A set-theoretic semantics for \(n\)-logic becomes interesting only when we have in our possession an independent (and independently interesting) description of homotopy \(n\)-types. Such is certainly the case for \(n = 1\) where groupoids admit a very simple description as categories all of whose morphisms are invertible. We will now give such a semantics by extending Makkai’s
original functorial semantics for FOLDS. Throughout, we will assume that we are working in a classical set theory, i.e. one in which subobject lattices are Boolean.

As before, fix a 1-signature \( \mathcal{L} \). Let \( V \) be a (strict, i.e. with on-the-nose preservation of identities and compositions) functor as in Definition 5.1.7 except we now postcompose it with the inclusion \( \text{Set} \hookrightarrow \text{Gpd} \) so that view it as a functor

\[
V : \mathcal{L} \to \text{Gpd}
\]

where \( \text{Gpd} \) is the 2-category of groupoids viewed as a full subcategory of the 2-category of categories \( \text{Cat} \). As before, a context is a finite (necessarily strict) subfunctor \( \Gamma \) of \( V \) and a context morphism a (necessarily strict) natural transformation between such subfunctors.

**Definition 5.4.1 (Set-theoretic \( \mathcal{L} \)-structure).** A (set-theoretic) \( \mathcal{L} \)-1-structure \( \mathcal{M} \) is a pseudofunctor \( (\mathcal{M}, \mu) : \mathcal{L} \to \text{Gpd} \) such that:

1. For each \( K \) with \( d(K) = 1 \) we fix

\[
\mathcal{M}(\equiv^1_K) \equiv \text{mor}\mathcal{M}(K)
\]

where we regard \( \text{mor}\mathcal{M}(K) \) as the arrow groupoid \( \mathcal{M}(K)^{\bullet \rightarrow \bullet} \) and interpret \( \mathcal{M}(s^1_K) \) and \( \mathcal{M}(t^1_K) \) as the domain and codomain functors (indeed, fibrations)

2. For each \( K \) with \( d(K) = 0 \) we interpret \( \equiv^1_K \) as equality of objects of the groupoid \( \mathcal{M}(K) \), i.e. fix

\[
\mathcal{M}(\equiv^1_K) \equiv \{(a, a) | a \in \text{ob}\mathcal{M}(K)\}
\]

regarded as a discrete groupoid and with \( \mathcal{M}(s^1_K) \) and \( \mathcal{M}(t^1_K) \) interpreted as the obvious projections.

273
(3) We fix

\[ \mathcal{M}(r_K^1) \equiv \{1_a | a \in \text{ob}\mathcal{M}(K)\} \]

regarded as a discrete groupoid and with \( \mathcal{M}(\rho_K^1) \) the obvious inclusion.

**Remark 5.4.2.** Clearly, not every functor \( F : \mathcal{L} \to \mathbf{Gpd} \) is an \( \mathcal{L} \)-structure since we require fixed denotations for logical sorts. Indeed, the variable functor \( V \) is not an \( \mathcal{L} \)-structure anymore as in Makkai's original definition, nor is any context \( \Gamma \). Thus, we may no longer view both the syntax and the semantics of \( n \)-logic as constructed functorially in exactly the same way. This asymmetry between syntax and semantics is inevitable when we move to \( n \)-logic for \( n \geq 1 \). This is because we want variables (as syntactic entities) to have strict identity conditions, i.e. we want them to be equal or unequal (and decidably so). Yet in \( n \)-logic such variables may denote entities with much coarser identity conditions, e.g. objects in a groupoid. Another way to think about this point is that if \( p \) is variable of sort \( x =_K^1 y \) (i.e. \( p \in V(=K^1) \) and \( V(s_K^1)(p) = x \) and \( V(t_K^1)(p) = y \) then \( p \) does not denote an actual isomorphism between \( x \) and \( y \) because \( x \) and \( y \) are assumed distinct and so there should be no isomorphism (or equality) between them in \( V(K) \).

Recall that a **pseudonatural transformation** \( (\alpha, \eta) \) between pseudofunctors

\[ (F, \phi), (G, \psi) : \mathcal{K}_1 \to \mathcal{K}_2 \]

is a natural transformation in which naturality (i.e. commutation of the required squares) holds only up to coherent isomorphisms given by \( \eta \). (For a precise definition cf. B1.1.2 [73].) We now want to consider evaluations as pseudonatural transformations from contexts to \( \mathcal{L} \)-structures. In our situation the definition is simplified significantly by the fact that the codomain of our (pseudo)functors will always be a 1-category and the codomain functor takes values in discrete categories.
**Definition 5.4.3** (Evaluation). Given a context $\Gamma$ and an $\mathcal{L}$-1-structure $\mathcal{M}$, an *evaluation of $\Gamma$ in $\mathcal{M}$* is a pseudonatural transformation

$$(\alpha, \eta): \Gamma \Rightarrow (\mathcal{M}, \mu): \mathcal{L} \to \text{Gpd}$$

Explicitly an evaluation is given by the following data:

1. A collection of maps $\{\alpha_K: \Gamma(K) \to \mathcal{M}(K)\}_{K \in \text{ob}\mathcal{L}}$
2. For every $f: K \to K_f$ in $\mathcal{L}$ a natural isomorphism
   $$\eta_f: \mathcal{M}(f) \circ \alpha_K \to \alpha_{K_f} \circ \Gamma(f)$$

   (where $\circ$ denotes horizontal composition of 2-cells) such that for any $g: K_f \to K_{gf}$ we have
   $$\eta_{gf} = \eta_g \circ g(\eta_f) \circ \mu_{f,g}$$

We will write $\text{pNat}(\Gamma, \mathcal{M})$ for the set of pseudonatural transformations between pseudofunctors $\Gamma$ and $\mathcal{M}$.

**Remark 5.4.4.** The “pseudo” part for natural transformations is essential: without it our semantics will not be sound with respect to the proof system developed in Section 5.5. On the other hand, the “pseudo” can be dropped from the definition of $\mathcal{L}$-structures in the sense that completeness can be proved with respect to the class of strict $\mathcal{L}$-structures (i.e. those defined by strict functors $\mathcal{L} \to \text{Gpd}$). Nevertheless, our semantics at this point can only gain from the added generality, and so we will not impose this restriction on ourselves.

Let us write $y$ for the Yoneda embedding post-composed with the inclusion $\text{Set} \hookrightarrow \text{Gpd}$ and $\hat{y}$ for the subfunctor of $y$ that misses the identity on the given object, i.e.
$\hat{y}X(X) = yX(X) \setminus \{1_X\}$. Given a variable $x \in |\Gamma|$ we can define $\partial x : \hat{y}K \Rightarrow \Gamma$ by $f \mapsto \Gamma(f)(x)$. Given any evaluation $(\alpha, \eta)$ and any variable $x : K$ we can define the $\alpha$-boundary of $x$ to be the composite $\alpha \circ \partial x$, and we write it as $\partial_\alpha x$. (Note that since $\partial x$ is a strict natural transformation between strict functors, $\alpha \circ \partial x$ become a pseudonatural transformation simply by inheriting $\eta$.) We also write $\alpha \ast a$ for the class of maps (which is not necessarily a (pseudo)natural transformation) that takes $x$ to $a$ and restricts to $\alpha$ otherwise.

The key difference from standard semantics of FOLDS is in how we define the $x$-range of $\alpha$. Intuitively we want this to be the set of all those terms in $\mathcal{M}$ that could serve as interpretations of the variable $x$ in the given evaluation $\alpha$. For this to make sense we must require that these terms $a \in \text{ob}\mathcal{M}(K)$ are “supported” by coherent isomorphisms that can extend $\alpha$ into a pseudonatural transformation that takes $x$ to $a$. This is the idea behind the following definition, which we can think of as the $\eta$-coherent $x$-range of $\alpha$.

**Definition 5.4.5 (x-range of an evaluation).** Let $(\alpha, \eta)$ be an evaluation of $\Gamma$ in $\mathcal{M}$ and $x \in \Gamma^\uparrow$. Then the $x$-range of $(\alpha, \eta)$ is defined as the following set

$$
\mathcal{M}_{(\alpha, \eta)}[x : K] = \{a \in \text{ob}\mathcal{M}(K), \bar{\eta} = (\bar{\eta}_f : \mathcal{M}(f)(a) \rightarrow (\partial_\alpha x)_K, f : K \rightarrow K)_{f \in \text{K}}

(\alpha \ast a, \eta \ast \bar{\eta}) \in \text{pNat}([\{\Gamma, x : K\}, \mathcal{M}])
$$

where by $\eta \ast \bar{\eta}$ we mean the amalgamation of the two coherence data into a single family of maps. (One can thing of $\ast$ simply as set-theoretic union if one takes this data to be given in the form of a set of arrows.) If $(a, \bar{\eta}) \in \mathcal{M}_\alpha[x]$, then we will write $\alpha[a/x]$ for the pseudonatural transformation that restricts to $\alpha$ in $\Gamma$ and takes $x$ to $a$, omitting the coherence data (as long as we know it exists, which might not always be the case).
Definition 5.4.6 (Extension of a formula). We define recursively the sets that comprise the extension $\mathcal{M}(\Gamma \mid \phi)$ of any $\mathcal{L}$-formula $\phi$ in context $\Gamma$ interpreted in $\mathcal{M}$ omitting explicit mention of coherence data:

\[
\mathcal{M}(\Gamma \mid \top) = \text{def } p\text{Nat}(\Gamma, \mathcal{M})
\]
\[
\mathcal{M}(\Gamma \mid \bot) = \text{def } \emptyset
\]
\[
\mathcal{M}(\Gamma \cup \Delta \mid \phi \land \psi) = \text{def } \mathcal{M}(\Gamma \mid \phi) \cap \mathcal{M}(\Delta \mid \psi)
\]
\[
\mathcal{M}(\Gamma \cup \Delta \mid \phi \lor \psi) = \text{def } \mathcal{M}(\Gamma \mid \phi) \cup \mathcal{M}(\Delta \mid \psi)
\]
\[
\mathcal{M}(\Gamma \cup \Delta \mid \phi \rightarrow \psi) = \text{def } \{ \alpha \in p\text{Nat}(\Gamma, \mathcal{M}) \mid \text{If } \alpha \in \mathcal{M}(\Gamma \mid \phi) \text{ then } \alpha \in \mathcal{M}(\Gamma \mid \psi) \}
\]
\[
\mathcal{M}(\Gamma \mid \forall x \phi) = \text{def } \{ \alpha \in \mathcal{M}(\Gamma \mid \phi) \mid \forall a \in \mathcal{M}_a[x], \alpha[a/x] \in \mathcal{M}(\Gamma \cup \{x\} \mid \phi) \}
\]
\[
\mathcal{M}(\Gamma \mid \exists x \phi) = \text{def } \{ \alpha \in \mathcal{M}(\Gamma \mid \phi) \mid \exists a \in \mathcal{M}_a[x], \alpha[a/x] \in \mathcal{M}(\Gamma \cup \{x\} \mid \phi) \}
\]

For the quantifiers we have assumed that $x$ is a variable of sort $K$ free in $\phi$ and that $x \in \text{FV}(\phi)^{\uparrow}$.

Remark 5.4.7. The fact that we consider $p\text{Nat}(\Gamma, \mathcal{M})$ only as a set (ignoring extra structure that we may naturally put on it) is also essentially what makes our approach proof-irrelevant: we don’t care about how the extension of a formula can be embedded in the extension of another formula, but merely whether it can.

Remark 5.4.8. Note that in the form presented in Definition 5.4.6 our semantics will always validate the LEM since we have defined it only in (a material and classical) set theory. In order to generalize the interpretation to constructive models (e.g. a Kripke-style semantics) we need to define it, as usual, for suitably-structured categories with
non-Boolean subobject lattices. However, the equality sorts here make such a general
definition more involved than usual.

Definition 5.4.9 (Satisfaction). For any $\mathcal{L}$-structure $\mathcal{M}$ and $\phi$ any $\mathcal{L}$-formula we can define satisfaction as follows:

$$\mathcal{M} \models \phi[\alpha/\Gamma] \Leftrightarrow \alpha \in M(\Gamma | \phi)$$

Similarly we define satisfaction of sequents as follows:

$$\mathcal{M} \models \Gamma | \phi \vdash \psi \Leftrightarrow M(\Gamma | \phi) \subseteq M(\Gamma | \psi)$$

Since an evaluation $\alpha$ already contains information about its domain $\Gamma$, we can simply write $\phi[\alpha]$ for the evaluation of $\phi$ at $\alpha$. As one would expect, given any formula $\phi$ it always suffices to consider $\text{var}(\phi)$ as its context. When $\text{var}(\phi) = \emptyset$ we write $\mathcal{M} \models \phi$ if the unique pseudonatural transformation $!: \emptyset \Rightarrow \mathcal{M}$ is in $\mathcal{M}(\emptyset: \phi)$.

The usual substitution lemma ensuring that satisfaction is preserved under composition of evaluations (i.e. repeated substitutions) can be proved straightforwardly, noting the minor subtlety that we are composing pseudonatural transformations and not strict ones. We state it for the record, leaving the proof to the reader.

Lemma 5.4.10 (Substitution Lemma). Let $\Gamma, \Delta$ be contexts, $\delta: \Delta \Rightarrow \Gamma$ a context morphism and $\phi$ a formula in context $\Delta$. If $\alpha \in M(\Gamma: \delta(\phi))$ then $\alpha \circ \delta \in M(\Gamma: \phi)$. In other words:

$$\mathcal{M} \models \delta(\phi)[\alpha] \Rightarrow \mathcal{M} \models \phi[\alpha \circ \delta]$$

Given any two pseudonatural transformations $(\alpha, \eta), (\beta, \theta)$ a modification

$$\nu: (\alpha, \eta) \rightarrow (\beta, \theta)$$

278
is given by maps $\nu_K : \alpha_K \to \beta_K$ satisfying evident coherence conditions with respect to $\eta$ and $\theta$. Since the elements of the extensions of formulas in our semantics are pseudonatural transformations, we can consider modifications between. In particular, we will say that two evaluations $\alpha, \beta \in \mathcal{M}(\Gamma : \phi)$ are isomorphic, and write $\alpha \cong \beta$, if there is a (necessarily invertible) modification $\nu : (\alpha, \eta) \to (\beta, \theta)$. Explicitly, in our simplified setting, such an isomorphism is given by the following data:

1. For each $K \in \text{ob}\mathcal{L}$ and $x \in \Gamma(K)$ an arrow $(\nu_K)_x : \alpha_K(x) \to \beta_K(x)$ in $\mathcal{M}(K)$ such that
2. for every $f : K \to K_f$ in $\mathcal{L}$ we have

$$((\nu_{K_f})_{\Gamma(f)(x)}) \circ (\eta_f)_x = (\theta_f)_x \circ \mathcal{M}(f)((\nu_K)_x)$$

The following lemma establishes that our semantics does not distinguish between isomorphic evaluations.

**Lemma 5.4.11 ("Invariance under modifications").** Let $\mathcal{M} \models \phi[\alpha]$ and $\beta \cong \alpha$. Then $\mathcal{M} \models \phi[\beta]$.

**Proof.** We proceed by induction on complexity. The cases of $\top, \bot, \lor, \land$ and $\to$ follow immediately. We do the case of existential quantification. Let $(\alpha, \eta) \in \mathcal{M}(\Gamma | \exists x : K.\phi)$ and assume we are given $\nu : (\alpha, \eta) \to (\beta, \theta)$. This means there exist coherence data $(a, \bar{\eta}) \in \mathcal{M}(a, \eta)[x : K]$ such that $(\alpha[a/x], \eta \ast \bar{\eta}) \in \mathcal{M}(\Gamma, x : K | \phi)$. For any arrow $f : K \to K_f$ in $\mathcal{L}$ define

$$\bar{\theta}_f = \nu_{K_f} \circ \bar{\eta}_f$$

It is easy to check that this defines coherence data such that $(\beta[a/x], \theta \ast \bar{\theta})$ is a pseudonatural transformation. Now define $\tilde{\nu} : (\alpha[a/x], \eta \ast \bar{\eta}) \to (\beta[a/x], \theta \ast \bar{\theta})$ to be the modification given by $\tilde{\nu}_K = 1_a$ and $\tilde{\nu}_{K'} = \nu_{K'}$ for all $K' \neq K$. That $\tilde{\nu}$ is a modification
follows immediately from the fact that \( \nu \) is. By the inductive hypothesis we get that 
\[(\beta[a/x], \theta \ast \bar{\theta}) \in \mathcal{M}(\Gamma, x: K \mid \phi).\] 
By taking \((a, \bar{\theta})\) to be the required data in the \(\theta\)-
coherent \(\beta\)-range of \(x\) it follows that \((\beta, \theta) \in \mathcal{M}(\Gamma \mid \exists x: K.\phi)\). The case of universal quantification follows exactly analogously. \(\square\)

**Remark 5.4.12.** Lemma 5.4.11 licenses us to consider only equivalence classes of pseudonatural transformations up to modification as the extensions of our formulas. Indeed, for the purposes of this Chapter we are free to do so. (But note that taking such equivalence classes becomes more difficult in a type-theoretic metatheory.)

We conclude this section with a crucial example, that may allay a nagging suspicion: since all arrows in the interpretation of equality sorts are isomorphic to the identity, are we even able to assert the existence of non-trivial automorphisms? (This is a crucial design constraint for our system; for if we cannot even express that equality sorts can have non-reflexivity inhabitants, then nothing justifies the kind of groupoid interpretation we are here considering.)

**Example 5.4.13.** Consider the 1-signature \(\mathcal{L}\) with only one non-logical sort \(O\) of dimension 1. Define the (strict) \(\mathcal{L}\)-structure that takes \(O\) to \(\mathbb{Z}/2\), with the latter regarded as a groupoid with a single object \(a\) and a single non-identity arrow \(p: a \to a\). We want to show that 
\[
\mathbb{Z}/2 \models x: O \mid \exists q: x =^O_1 x. \forall \tau: r^O_1(q, x).\bot
\]
To do so, it suffices to show that

\[
\mathbb{Z}/2(\{x: O\} \mid \exists q: x =^O_1 x. \forall \tau: r^O_1(q, x).\bot)
\]
is non-empty. Write $\alpha$ for the obvious pseudonatural transformation $\{x: O\} \to \mathbb{Z}/2$ that sends $x$ to $a$. We now have:

$$\mathbb{Z}/2(\{x: O\} \mid \exists q: x = 1^0 x. \exists \tau: r^1_O(q, x). \bot) =$$

$$= \{\alpha | \exists (\bar{q}, \bar{\eta}) \in \mathbb{Z}/2[\eta] \text{ s.t. } a[\bar{q}/\eta] \in \mathbb{Z}/2(\{x, q\} : \forall \tau: r^1_O(q, x). \bot)\}$$

$$= \{\alpha | \exists (\bar{q}, \bar{\eta}) \forall (\bar{\tau}, \bar{\xi}) \in \mathbb{Z}/2^2[\eta][\tau] \text{ s.t. } \alpha[\bar{q}/\eta][\bar{\tau}/\tau] \in \emptyset\}$$

The last line above indicates that the required set is non-empty if there is a particular choice of $\bar{q}$ and $\bar{\eta}$ that cannot be coherently extended to a pseudonatural transformation. So it suffices to show that there is indeed such a choice of $\bar{q}$ and $\bar{\eta}$. To see this take

$$\bar{q} = \bar{\eta}_t = \bar{\eta}_s = p$$

where $t$ and $s$ denote the target and source morphisms in $L$ from $=^1_O$ to $O$. Similarly let $\rho$ denote the unique morphism $r^1_O \to =^1_O$. Now assume that there is an extension of this data into a pseudonatural transformation $(\alpha[\bar{q}/\eta][\bar{\tau}/\tau], \xi)$. The only possible choices for $\xi_\rho$ are $(p, 1_a)$ and $(1_a, p)$. But since $t \rho = s \rho$ we must have $\xi_{s \rho} = \xi_{t \rho}$ which means that

$$\eta_t \circ t(\xi_\rho) = \eta_s \circ s(\xi_\rho)$$

So for any of the two choices of $\xi_\rho$ we get $p \circ p = p$ which is a contradiction since $p$ is its own inverse in $\mathbb{Z}/2$.

**Remark 5.4.14.** We see in the above example that falsity in our system is best understood as incoherence. In particular, the existence of non-identity (auto)morphisms in Example 5.4.13 is parsed as the non-existence of a coherent extension of a certain choice of data. So although every morphism will be isomorphic to the identity, not all
of them will be coherently so. Our semantics is thus best understood as a semantics of *coherence*, not of *truth*. The two notions coincide for 0-logic but come apart for *n*-logic with *n* ≥ 1. Satisfaction for *n*-logic is not to be understood merely as the existence of certain data, but as the *coherent* existence of certain data.

5.5. Proof System for *n*-logic

We now describe a proof system *D* for *n*-logic. Since we want to work with possibly empty sorts as denotations (i.e. we do not want the inhabitation of sorts to be valid in our system) we have to “stratify” our formulas with respect to given contexts. The most convenient way of achieving this proof-theoretically is via a sequent calculus. On the other hand, as we shall see, our version of the *J*-rule forces us into at least the regular fragment of first-order logic (i.e. formulas built out of the connectives ∃ and ∧). Since conjunction will therefore always be available to us, this justifies our decision to restrict ourselves to “singular” sequents.

Let *L* be an *n*-signature. Then *D* consists of the standard rules of (intuitionistic or classical, coherent or full) first-order logic to which we add three rules. For the sake of completeness, we will lay out the “old” rules of *D* in their entirety. Our presentation combines elements from [71, 73, 110]. Clearly, for each fragment of first-order logic we wish to consider we employ only the rules corresponding to the connectives available in that fragment. As usual, the double lines denote a rule that goes in either direction and we assume that every formula that appears below is a well-formed *L*-formula and that every sequent is well-formed.

**Structural Rules**
(iden)  \[ \Gamma \vdash \phi \rightarrow \phi \]

(Sub)  \[ \Delta \vdash s(\phi) \rightarrow s(\psi) \]
\[ \Gamma \vdash \phi \rightarrow \psi \]

(Cut)  \[ \Gamma \vdash \phi \rightarrow \psi \]
\[ \Gamma \vdash \psi \rightarrow \chi \]
\[ \Gamma \vdash \phi \rightarrow \chi \]

(Con-wk)  \[ \Gamma \vdash \phi \rightarrow \psi \]
\[ \Gamma, x: K \vdash \phi \rightarrow \psi \]

(Con-exch)  \[ \Gamma, y: K', x: K, \Gamma' \vdash \phi \rightarrow \psi \]
\[ \Gamma, x: K, y: K', \Gamma' \vdash \phi \rightarrow \psi \]
\[ x \notin \text{dep}(y) \]

Logical Rules

(⊤)  \[ \Gamma \vdash \top \]

(⊥)  \[ \Gamma \vdash \bot \]

(∧)  \[ \Gamma \vdash \theta \rightarrow \phi \]
\[ \Gamma \vdash \theta \rightarrow \psi \]
\[ \Gamma \vdash \theta \rightarrow \psi \rightarrow \phi \rightarrow \psi \]

(V)  \[ \Gamma \vdash \phi \rightarrow \theta \]
\[ \Gamma \vdash \psi \rightarrow \theta \]
\[ \Gamma \vdash \phi \lor \psi \rightarrow \theta \]

(→)  \[ \Gamma \vdash \theta \rightarrow \phi \rightarrow \psi \]

(∀)  \[ \Gamma, x: K \vdash \theta \rightarrow \phi \]
\[ x \notin \text{var}(\Gamma) \]

(∃)  \[ \Gamma, x: K \vdash \theta \rightarrow \phi \]
\[ x \notin \text{var}(\Gamma) \]

If we are working in the regular or coherent fragment then we also add the following rule, which is otherwise derivable:

(Frob)  \[ \Gamma \vdash \phi \rightarrow (\exists x: K.\psi) \]
\[ x \notin \text{var}(\phi) \]
\[ \Gamma \vdash \phi \rightarrow (\exists x: K.\psi) \rightarrow \forall x: K.\phi \rightarrow \psi \]
If we are working in the coherent fragment then we add the following rule, which is otherwise derivable:

\[ \Gamma | \phi \land (\psi \lor \chi) \vdash (\phi \land \psi) \lor (\phi \land \chi) \]  

(Dist)

Finally, to get \( D^{cl} \) we can add the law of the excluded middle:

\[ \Gamma | \top \vdash \phi \lor (\phi \to \bot) \]  

(LEM)

Now we present the “new” rules. The first new rule corresponds to the introduction rule for identity types in MLTT:

\[ \Gamma, x : K | \theta \vdash \exists p : x =^1_K x. \exists r : r^1_K(p, x). \top \]  

(Eq-intro)

The second rule is the crucial J-rule, which corresponds to the elimination rule for identity types in MLTT:

\[ \Gamma, x : K | \exists q_1 : x =^1_K x. r^1_K(q_1) \land \theta[x/y, q_1/p] \vdash \exists q_2 : x =^1_K x. r^1_K(q_2) \land \phi[x/y, q_2/p] \]  

(J)

**Remark 5.5.1.** This is where the decision to include reflexivity sorts of dimension \(-2\) pays off. It allows us to have a uniform presentation of the rules for equality sorts without splitting them into “propositional” and “non-propositional”, which would force us into two separate sets of rules (one for “top-level” equalities and one for the rest).
What we can do instead is prove that “top-level” equalities behave like propositions, as indeed we shall do in Proposition 5.5.3 below.

Finally, we add the following rule expressing the fact that “reflexivity is unique up to equality one level up”:

\[
\Gamma, x: K, p: x =_K x, q: x =_K x \mid r_K^1(p) \land r_K^1(q) \vdash \exists \alpha: p =_K q. \top
\]

(R)

Remark 5.5.2. In our presentation above we have assumed that every sequent is well-formed and that the rules are only instantiated for appropriate sorts (e.g. \(\text{Eq-intro}\) is instantiated only for \(K\) with \(d(K) < -1\)). Furthermore, we have suppressed all variable dependencies that are not directly relevant to the rule in question and have also assumed that sorts \(K\) as they appear could themselves denote equality sorts (in which case we understand \(=_K^1\) as an abbreviation for \(=_{K'}^{i+1}\) where \(K = _{K'}^i\)). Finally, the notation \(\phi[x/y, q/p]\) in \((J)\) denotes the formula \(\delta(\phi)\) obtained by substitution along the contraction morphism described in Example 5.1.10.

If we so choose we can certainly add the law of the excluded middle (LEM) as an axiom. When we do so we will denote the corresponding proof system by \(D^{cl}\). When what we say applies to both \(D\) and \(D^{cl}\) we will use the notation \(D^{(cl)}\).

We note that the \((J)\) rule above corresponds to what in MLTT would be called strong \(\text{Id}\)-elimination since \(\theta\) behaves like a contextual parameter that may itself depend on the variables \(\{x, y, p\}\). In the presence of \(\Pi\)-types, strong \(\text{Id}\)-elimination is equivalent to the usual form (without a contextual parameter). Similarly, in the presence of universal quantification \((J)\) is equivalent to the more recognizable rule.
The \( (J) \)-rule can be used to prove that “top-level” equality behaves exactly like a proposition. This fact means that the “top-level” equality sort of any other sort \( K \in \mathcal{L} \) (of non-maximal level) will behave like a proposition, the “second-from-top” equality sort will behave like a set etc. Our proof system thus captures the crucial element that our (homotopy) semantics demands: that sorts of dimension \( m \) behave like \( m \)-types. The following proposition registers this fact.

**Proposition 5.5.3 (“Top-level equality is propositional”).** For any \( K \in ob\mathcal{L} \) with dimension \( d(K) = 0 \) and any \( \mathcal{L} \)-formula \( \phi \) in context \( \Gamma, x: K, y: K \) the following rule (called “Lawvere equality” in [71]) is derivable:

\[
\begin{align*}
\Gamma, x: K & \mid \theta \vdash \phi[x/y] \\
\Gamma, x: K, y: K & \mid \theta, \exists p: x =^K y \vdash \phi
\end{align*}
\]

**Proof.** Since \( K \) is assumed to be of dimension 0, we have \( d(-^K) = -1 \) and therefore for any formula \( \phi \) the substitution \( \phi[x/y] \) is well-defined. We then have:

\[
\begin{align*}
\Gamma, x: K & \mid \theta \vdash \phi[x/y] \\
\Gamma, x: K & \mid \theta \vdash \phi[x/y] \land \exists q: x =^K x.r^1_K(q) \\
\Gamma, x: K & \mid \theta \vdash \exists q: x =^K x.(\phi[x/y] \land r^1_K(q)) \\
\Gamma, x: K,y: K & \mid \theta, \exists p: x =^K y \vdash \phi
\end{align*}
\]
In applying (∃-intro) and (Frob) we have assumed that p does not appear in φ, as indeed we can do without loss of generality.

One gets transport for free in HoTT (cf. [171], p.91) as a consequence of the J-rule for identity types of the ambient type theory. By a similar argument, we also get the relevant version of transport in our setting.

**Proposition 5.5.4 (“Transport”).** For any $\mathcal{L}$-formula $\phi$ in context $\{\Gamma, x : K\}$ the sequent

$$\Gamma, x : K, y : K, p : x =_K y \mid \phi \vdash \phi[y/x]$$

is derivable.

**Proof.** By (iden) we know that the sequent

$$\Gamma, x : K \mid \phi[x/y] \vdash \phi[y/x][x/y]$$

is derivable (from no premises) since $\phi[x/y] \equiv \phi[y/x][x/y]$. But then we can just repeat the (first three steps of) the derivation in the proof of Proposition 5.5.3 with $\theta \equiv \phi[x/y]$ and $\phi \equiv \phi[y/x]$ to get the desired result. □

We may now define the notion of $n$-theory in the usual manner: an $\mathcal{L}$-$n$-theory is given by (the $D^{(cl)}$-closure of) a set of $\mathcal{L}$-$n$-sentences (its axioms). As usual we will write $\mathcal{M} \models \mathbb{T}$ for an $\mathcal{L}$-$n$-structure $\mathcal{M}$ that satisfies all the axioms of $\mathbb{T}$. Following [171], we write LEM for the following type in UF:

$$\Pi_{A : \text{Prop}_{cl}} A + \neg A$$

For any $\mathcal{L}$-$n$-theory $\mathbb{T}$ we write $\mathbb{T} \vdash \phi$ (resp. $\mathbb{T} \vdash_{cl} \phi$) to denote that $\phi$ is derivable in $D$ (resp. $D^{(cl)}$) from $\mathbb{T}$. We write $\mathcal{M} \models \phi$ (resp. $\mathcal{M} \models_{cl} \phi$) to denote that $\phi$ is true in all models.
of $T$ in HoTT (resp. HoTT+LEM) and similarly for the set-theoretic semantics. At this point we have obtained all the components traditionally required of a logic: a syntax, a semantics and a proof system. We may thus begin to investigate the relationship between these components, as indeed we do in the next section.

**Remark 5.5.5.** One might wonder whether the class of (homotopy) models we are considering for $n$-logic is too wide. In particular, since we appear not to be making any use of the full structure of identity types one might wonder whether our $n$-logic can take semantics where *every* type is interpreted as a set (i.e. a 0-type) much as in Makkai’s original formulation of the semantics of FOLDS. This is not the case. $n$-logic does in fact have the expressive power to force a theory to have only models whose ground sorts are $n$-types. As the simplest possible illustration, take the case where $n = 1$ and $L$ is the same signature as in Example 5.4.13. Consider the $L$-1-theory $T_O$ consisting of the single axiom

$$\phi \equiv \forall x, y : O. \exists p, q : x =^1_O y. \neg (p =^2_O q)$$

with the obvious abbreviations. Every model of $T_O$ where $O$ is interpreted as an $h$-set (resp. discrete groupoid) falsifies $\phi$. But $T_O$ is satisfiable: simply take an $L$-1-structure where $O$ is interpreted as a (proper) 1-type (resp. groupoid). Assuming the soundness results of Section 5.6 this means that $\phi$ cannot be disproved by $D$ even though it is not satisfied in any set-model of $T_O$. As such, set models are not sufficient to describe provability for 1-logic. Exactly analogously, we can see that $(n - 1)$-type models are not sufficient to describe provability for $n$-logic.

**Remark 5.5.6.** Even if $n$-logic is expressive enough to go beyond set-theoretic semantics, one may still suspect that $D^{(cl)}$ is much too impoverished to capture enough of the complexity of identity types in MLTT to prove a completeness theorem. Recall
however, that the syntax of $n$-logic does not include any closed terms and the actual equality sorts have already been introduced through the process of globular completion. Therefore, the rules for identity types whose analogues do not appear in $D^{(cl)}$ (i.e. $\text{Id}$-intro and $\text{Id}$-comp) would not sensibly translate to anything that we may wish to do in $D$. ($\text{Id}$-intro has already been applied in the formation of formulas and $\text{Id}$-comp applies to closed terms which we do not have). This should give some initial plausibility that the (homotopy) semantics we have outlined are complete with respect to $D$ since there is nothing else in MLTT that one can do with $h$-props that cannot be done with the (suitable translates of the) rules of $D$.

5.6. Soundness

We will now prove soundness for the rules of $D$ with respect to both our semantics. We begin with the homotopy semantics of Section 5.3.

**Theorem 5.6.1 (Soundness for homotopy semantics).** If $T \vdash_{(cl)} \phi$ then $T \models_{(cl)} \phi$.

**Proof.** All the traditional rules of $D$ have direct analogues in MLTT and so the proof proceeds without difficulties by induction on the complexity of derivations. This means that we take each particular rule in the deductive system and show that there is a derivation from the translation of the top line to the translation of the bottom line. There is one minor subtlety. Axioms in $D$ are stated by starting from an empty line and then producing a formula. For example, in $D$ we have the following axiom

$$
\begin{array}{c}
\Gamma | \phi \vdash \phi \\
\end{array}
$$

iden

There is a certain amount of information suppressed in stating such an axiom, namely that the given sequent is well-formed. In translating an axiom like (iden) we will therefore...
translate the “empty” set of formulas above as the judgement in HoTT stating that the
types involved in the translation of the formula below are well-formed. (This is essentially
the price we pay when we interpret a proof-irrelevant system into a proof-relevant one.) So
in translating (iden) we need to show that there is a HoTT-derivation from the judgment

\[ \Delta^M, \Gamma^M \vdash \phi^M : U \]

to the judgment

\[ \Delta^M \vdash s : \Pi_{x : \Gamma^M} \phi^M \rightarrow \phi^M \]

for some s and for any M. In a somewhat abbreviated form this goes as follows

\[
\begin{array}{c}
\Delta^M, \Gamma^M \vdash \phi^M : U \\
\Pi\text{-form, wkg}
\hline
\Delta^M, \Gamma^M \vdash \phi^M : U \\
\Pi\text{-form}
\hline
\Delta^M \vdash \Pi_{x : \Gamma^M} \phi^M \rightarrow \phi^M : U \\
\Pi\text{-intro}
\end{array}
\]

where the term \( \lambda \vec{x}.(\lambda y.y) \) has been produced by applying \( \Pi\text{-intro} \) to

\[ \Delta^M, \Gamma^M, y : \phi^M \vdash y : \phi^M \]

We do the same for other axioms. Since the languages we are considering are purely
relational and there are no closed terms, the substitution rule follows immediately without
any complications. As for the cut rule, given terms

\[ \Delta^M \vdash \eta : \Pi_{x : \Gamma^M} \phi^M \rightarrow \psi^M \]
and

\[ \Delta^M \vdash \xi : \Pi_{\bar{x} : \Gamma^M} \psi^M \rightarrow \chi^M \]

we can define a term

\[ \Delta^M \vdash \lambda \bar{x}.(\lambda y : \phi^M.\xi(\eta(\bar{x})(y)))(\eta(\bar{x}))(y)) : \Pi_{\bar{x} : \Gamma^M} \phi^M \rightarrow \chi^M \]

The rest of the structural rules follow just as easily.

The logical rules follow by interpreting the connectives in the manner of Section 5.3. We will do the case of existential quantification as an illustration. The relevant rule is the following

\[ \begin{array}{c}
\Gamma, x : K \vdash \phi \vdash \psi \\
\hline 
\Gamma \mid \exists x : K.\phi \vdash \psi
\end{array} \]

So suppose we have derived the judgement

\[ \Delta^M \vdash \eta : \Pi_{\bar{x} : \Gamma^M} \psi^M \rightarrow \psi^M \]

Then we can define the following term

\[ \Delta^M \vdash \xi \equiv \lambda \bar{x}.(\lambda (y, p).\eta(\bar{x}, y)(p)) : \Pi_{\bar{x} : \Gamma^M \ y : K^M} \Sigma_{\bar{x} : \Gamma^M \ y : K^M} \phi^M \rightarrow \psi^M \]

But since \( \psi^M \) will be an \( h \)-prop for any substitution instance of its free variables, by the universal property of the propositional truncation we get a term

\[ \Delta^M \vdash \xi \equiv \lambda \bar{x}.\text{||}(\lambda (y, p).\eta(\bar{x}, y)(p))\text{||} : \Pi_{\bar{x} : \Gamma^M \ y : K^M} \Sigma_{\bar{x} : \Gamma^M \ y : K^M} \phi^M\text{||} \rightarrow \psi^M \]

This is exactly the translation of the (satisfaction of the) bottom sequent in (\( \exists \)). The rest of the logical rules follow similarly, employing the universal property of the propositional truncation when needed.
So it remains to prove soundness for (Eq-intro), (J) and (R). For (Eq-intro) we have that

\[ \Delta^\mathcal{M} \vdash (\text{refl}_x, \text{refl}_{\text{refl}_x}) : \big|\big| \sum_{p : \text{Id}_{K^\mathcal{M}}(x,x)} (p, \text{refl}_x) \big|\big| \]

is derivable which gives us the required result. For (J), given the availability of \( \Pi \)-types in our metatheory, we will prove (J') for simplicity since the main idea of the argument is exactly the same. So suppose we are given:

\[ \mathcal{M} \models \Gamma, x : K \mid \top \vdash \exists q : x =_K x. \exists \tau : r_K(q). \phi[x/y, q/p] \]

This means that the following judgement is derivable in HoTT:

\[ \Delta^\mathcal{M} \vdash \eta : \prod_{\vec{z} : \Gamma^\mathcal{M}} \big|\big| \sum_{q : \text{Id}_{K^\mathcal{M}}(x,x)} (q, \text{refl}_x) \times \phi^\mathcal{M}[x/y, q/p] \big|\big| \]

Since \( \phi^\mathcal{M} \) is an \( h \)-prop we can remove it from the propositional truncation to obtain that the following judgment is derivable (where once again denote the term by \( \eta \)):

\[ \Delta^\mathcal{M} \vdash \eta : \prod_{\vec{z} : \Gamma^\mathcal{M}} \big|\big| \sum_{q : \text{Id}_{K^\mathcal{M}}(x,x)} (q, \text{refl}_x) \big|\big| \times \phi^\mathcal{M}[x/y, q/p] \]

Now we can define a term

\[ \xi : \prod_{\vec{z} : \Gamma^\mathcal{M}} \sum_{q : \text{Id}_{K^\mathcal{M}}(x,x)} (q, \text{refl}_x) \times \phi^\mathcal{M}[x/y, q/p] \rightarrow \phi^\mathcal{M}[x/y, q/\text{refl}_x] \]

by

\[ \xi \equiv \lambda \vec{z}. \lambda w. \pi_1(\text{pr}_2(w))^*(\pi_2(\text{pr}_2(w))) \]
and since \( \phi^M \) will always be an \( h \)-prop once again by the universal property of the propositional truncation this gives us a term

\[
\xi : \prod_{z : \Gamma^M} \| q : \text{Id}_{K^M(x,x)}(q, \text{refl}_x) \| \times \phi^M[x/y, q/p] \to \phi^M[x/y, q/\text{refl}_x]
\]

By using \( \eta \) we thus get a term

\[
\Delta^M \vdash \tilde{\eta} : \prod_{z : \Gamma^M} \phi^M[x/y, \text{refl}_x]
\]

But then by the induction rule for identity types (i.e. \( \text{Id} \)-elim internalized) we get

\[
\Delta^M \vdash \text{ind}_x(\tilde{\eta}) : \prod_{\bar{z} : (\Gamma \backslash \{y,p\})^M} \phi^M
\]

which gives us the desired result. For (R) we may simply apply the well-known transitivity of identity types, which immediately gives us the desired result. Finally, the soundness of LEM in HoTT+LEM is evident. \( \square \)

We now also prove soundness for the set-theoretic semantics outlined in Section 5.4. The main difficulty is once again in showing the soundness of the \((J)\)-rule. In the proof below this is achieved by arguing that the extension of any formula with a path in its free variables is fully (and coherently) determined by individual objects in the groupoid in which that path is interpreted.

**Theorem 5.6.2 (Soundness for set-theoretic semantics of 1-logic).** If \( \mathcal{T} \models_{(cl)} \phi \) then \( \mathcal{T} \models_{(cl)} \phi \).

**Proof.** The structural rules and the rules for the connectives follow immediately just as in normal first-order logic. (R) and (Eq-intro) also follow immediately since any sort of dimension 1 is interpreted as a groupoid and will by definition contain unique (up...
to equality) identity arrows. As before, the only non-trivial case is (J). We will consider
the case of (J') since the core of the argument is exactly the same and keeping track of
parameters unnecessarily obfuscates it. Furthermore, since the formulas to which (J')
meaningfully applies will depend on sorts of dimension 1 and hence will not depend on
other sorts below them, it suffices to consider the case where there is no ambient context
Γ. Therefore, it suffices to prove the soundness of the following simplified rule

\[
\begin{align*}
\text{x: } K & \mid \top \vdash \exists q: x =_K x.\exists r: r_1K(q, x).\phi[x/y, q/p] \\
\text{x: } K, y: K, p: x =_K y & \mid \top \vdash \phi
\end{align*}
\]

(J’)

where \(\phi[x/y, q/p]\) denotes the formula \(\delta(\phi)\) obtained from the change-of-variables \(\delta: \{x, y, p\} \Rightarrow \{x, q\}\) described in Example 5.1.10.

So suppose we are given an \(L\)-structure \(M\) such that

\[
\mathcal{M}(\{x: K\} \mid \top) \subseteq \mathcal{M}(\{x: K\} \mid \exists q: x =_K x.\exists r: r_1K(q, x).\phi[x/y, q/p])
\]

We need to show that (*) implies

\[
\mathcal{M}(\{x, y: K, p: x =_K y\} \mid \top) \subseteq \mathcal{M}(\{\Gamma, x, y: K, p: x =_K y\} \mid \phi)
\]

Assume for simplicity that \(L\) is strict. Take an arbitrary pseudonatural transformation
(\(\alpha, \eta\)): \(\{x, y, p\} \Rightarrow \mathcal{M}\) which consists of the following data

\[
a = \alpha_K(x) \quad b = \alpha_K(y)
\]

\[
\alpha_{x =_K p} = a' \xrightarrow{\bar{p}} b'
\]

\[
\eta_s: a' \rightarrow a \quad \eta_t: b' \rightarrow b
\]
where as usual $s$ and $t$ denote the source and target maps of the equality sort $=^1_K$.

Consider now the pseudonatural transformation $[a/x]: \{x\} \Rightarrow M$ that takes $x$ to $a$ (and contains trivial coherence data). By (*) we know that

\[(1) \quad [a/x] \in M(\{x: K\} | \exists q: x =^1_K x. \exists r: r^1_K (q, x) x =^1_K x. \exists \tau: r^1_K (q, x) x =^1_K x. \phi[x/y, q/p])\]

In particular this means that there exists $\bar{q}: \alpha'' \to \beta''$, $\theta_s: a'' \to a$, $\theta_t: b'' \to a$ such that, by Lemma 5.4.10, we get

\[(\beta, \theta) \equiv ([a/x][a/y][\bar{q}/p], \{\theta_s, \theta_t\}) \in M(\{x, y, p\} | \phi)\]

We now claim that $(\alpha, \eta) \simeq (\beta, \theta)$. To define a modification $\nu: (\alpha, \eta) \to (\beta, \theta)$ we need arrows

\[
(\nu_K)_x: \alpha_K(x) = a \to a = \beta_K(x)
\]

\[
(\nu_K)_y: \alpha_K(y) = b \to b = \beta_K(x)
\]

\[
(\nu_{-p}^1)_k: \alpha_{-p}^1 (p) = \bar{p} \to \bar{q} = \beta_{-p}^1 (p)
\]

satisfying the required coherence condition

\[
\theta_f \circ M(f)((\nu_{-p}^1)_k) = (\nu_K)_{\Gamma(f)(p)} \circ \eta_f
\]
where we have used $\Gamma$ as notation for the associated functor of the given context $\{x, y, p\}$ and where $f \in \{s_{K}^{1}, t_{K}^{1}\}$. We fill this data as follows:

\[
\begin{align*}
(\nu_{K})_{x} & \equiv \theta_{sp} \\
(\nu_{K})_{y} & \equiv \theta_{tp} \circ \eta_{s} \circ (\eta_{t} \circ \bar{p})^{-1} \\
(\nu_{K}^{-1})_{p} & \equiv ((\theta_{p})_{1} \circ \eta_{s}, (\theta_{p})_{2} \circ \eta_{s} \circ \bar{p}^{-1})
\end{align*}
\]

where $\theta_{p}$, $\theta_{tp}$ and $\theta_{sp}$ are given by the fact that we know by (1) that there exists a coherent isomorphism from $1_{a}$ to $\bar{q}$. It remains to check that $\nu$ satisfies the required coherence conditions. Firstly, for $s$, we have

\[
\begin{align*}
\theta_{s} \circ M(s)((\nu_{K}^{-1})_{p}) & = \theta_{s} \circ (\theta_{p})_{1} \circ \eta_{s} \\
& = \theta_{s} \circ M(s)(\theta_{p}) \circ \eta_{s} \\
& = \theta_{sp} \circ \eta_{s} \\
& = (\nu_{K})_{x} \circ \eta_{s}
\end{align*}
\]

Finally, for $t$, we have

\[
\begin{align*}
\theta_{t} \circ M(t)((\nu_{K}^{-1})_{p}) & = \theta_{t} \circ (\theta_{p})_{2} \circ \eta_{s} \circ \bar{p}^{-1} \\
& = \theta_{t} \circ M(t)(\theta_{p}) \circ \eta_{s} \circ \bar{p}^{-1} \\
& = \theta_{tp} \circ \eta_{s} \circ \bar{p}^{-1} \\
& = \theta_{tp} \circ \eta_{s} \circ (\eta_{t} \circ \bar{p})^{-1} \circ \eta_{t} \\
& = (\nu_{K})_{y} \circ \eta_{t}
\end{align*}
\]
This establishes that \( \nu \) is a modification and therefore that \( (\alpha, \eta) \cong (\beta, \theta) \). By Lemma 5.4.11 this means that \( (\alpha, \eta) \in \mathcal{M}(\{x, y, p\} \mid \phi) \) and therefore that

\[
\mathcal{M}(\{x, y, p\} \mid \phi) = p\text{Nat}(\{x, y, p\}, \mathcal{M})
\]

which is exactly \((**\))%. Finally, as noted in Remark 5.4.8, LEM is clearly sound since we are working in a classical set theory.

\[\square\]

### 5.7. Examples and Applications

FOLDS was invented as a systematic way of avoiding the use of equalities that are irrelevant for the structures of interest, e.g. equality between objects when we care about categories only up to equivalence. On the other hand, \( n \)-logic represents a reversal of this idea, since we are re-introducing equalities as logical sorts with a fixed interpretation. But we don’t have to do this in order to get some traction out of FOLDS as a proof-irrelevant syntax with a natural semantics in HoTT. Indeed, normal FOLDS can be employed to axiomatize structures in UF as long as we allow ourselves to fix the interpretation of certain sorts as identity types in HoTT. Another (equivalent) way to achieve this is simply to extend the dimension functions \( d \) to \( \mathbb{Z}_{\geq -2} \cup \{\infty\} \), i.e. to allow for certain sorts to be types of arbitrarily large dimension. More precisely, we will call \((\mathcal{L}, d)\) an \( \infty \)-signature if it is a FOLDS-signature where \( d \) is allowed to take values in \( \mathbb{Z}_{\geq -2} \cup \{\infty\} \) and where the interpretation of a sort \( K \) with \( d(K) = \infty \) is an arbitrary type \( K^M : \mathcal{U} \).
Now, let $\mathcal{L}_{\text{cat}}$ denote the following $\infty$-signature

subject to the relations

\[ d_0 t_0 = d_0 t_2, \quad d_1 t_1 = d_1 t_2, \quad d_0 t_1 = d_1 t_0 \]

\[ d_0 i = d_1 i \]

\[ d_0 e_1 = d_0 e_2, \quad d_1 e_1 = d_1 e_2 \]

The $\mathcal{L}_{\text{cat}}$-theory of categories $\mathbb{T}_{\text{cat}}$ consists of the following axioms:

(1) (Existence of identities)

\[ \forall x: O. \exists i: A(x, x). \exists \sigma: I(i, x, x). \top \]

(2) (Functionality of composition-1)

\[ \forall x, y, z: O. \forall f: A(x, y). \forall g: A(y, z). \exists h: A(x, z). \exists \tau: \circ (f, g, h, x, y, z). \top \]

(3) (Functionality of Composition-2)

\[ \forall x, y, z: O. \forall f: A(x, y). \forall g: A(y, z). \forall h, h': A(x, z). \forall \tau_1: \circ (f, g, h). \forall \tau_2: \circ (f, g, h'). \exists \epsilon: =_A (h, h', x, z) \]
(4) (Associativity)
\[ \forall x, y, z, w : O. \forall f : A(x, y). \forall g : A(y, z). \forall h : A(z, w). \forall i : A(x, z). \forall j : A(x, w). \]
\[ \forall k : A(y, w). \forall \tau_1 : o(f, g, i, x, y, z). \forall \tau_2 : o(i, h, j, x, z, w). \forall \tau_3 : o(g, h, k, y, z, w). \]
\[ \exists \tau_4 : o(f, k, j, x, y, w). \top \]

(5) (Uniqueness of identity)
\[ \forall x : O. \forall i, j : A(x, x). \forall \sigma_1 : I(i, x, x). \forall \sigma_2 : I(j, x, x). \exists \epsilon :: =_A (i, j, x, x). \top \]

(6) (Right unit)
\[ \forall x, y : O. \forall i : A(x, x). \forall g : A(x, y). \forall \sigma : I(i, x, x). \exists \tau : o(i, g, g, x, y). \top \]

(7) (Left unit)
\[ \forall x, y : O. \forall i : A(y, y). \forall f : A(x, y). \forall \phi : I(i, y, y). \exists \tau : o(f, i, f, x, y, y). \top \]

Using our homotopy semantics we get that an \( L_{\text{cat}} \)-structure consists of the following data:

- A type \( O : U \)
- A term \( A : O \to O \to \text{Set}_d \)
- A term \( I : \Pi_{a : O} A(x, x) \to \text{Prop}_d \)
- A term \( o : \Pi_{x,y,z : C} A(x, y) \to A(y, z) \to A(x, z) \to \text{Prop}_d \)

We can now translate the axioms of \( \mathcal{T}_{\text{cat}} \) into types in HoTT for an arbitrary model \( \mathcal{M} \) of \( \mathcal{T}_{\text{cat}} \). We will list them in order, writing \( = \) for the identity on \( A^\mathcal{M} \) and omitting \( \mathcal{M} \) from superscripts:

\[
\begin{align*}
(T_1) & \quad \Pi_{x : O} \|I(i, x, x)\| \\
(T_2) & \quad \Pi_{x,y,z : O} \| \sum_{f : A(x, y)} \sum_{g : A(y, z)} o(f, g, h, x, y, z)\|
\end{align*}
\]

\[
(T_3) \quad \Pi_{x,y,z : O} h = h'
\]

\[
\begin{align*}
\tau_1 & : o(f, g, h, x, y, z) \\
\tau_2 & : o(f, g, h', x, y, z)
\end{align*}
\]
We thus obtain:

\[ \text{Mod}(\mathbb{T}_{\text{cat}}) \equiv \sum_{A: O \to \text{Set}_{\mathcal{U}}} \prod_{a: C} \text{Hom}_{C}(a, a) \times \prod_{a, b: C} \text{Hom}_{C}(a, b) \times \text{Hom}_{C}(b, c) \times \text{Hom}_{C}(a, c) \]

Following [171], a \textit{precategory} is defined by the following data:

1. A type \( C: \mathcal{U} \) ("objects")
2. A dependent type \( \text{Hom}_{C}: C \to C \to \text{Set}_{\mathcal{U}} \) ("Hom-sets")
3. \( 1: \prod_{a: C} \text{Hom}_{C}(a, a) \)
4. \( \circ: \prod_{a, b, c: C} \text{Hom}_{C}(a, b) \to \text{Hom}_{C}(b, c) \to \text{Hom}_{C}(a, c) \)
5. \( \text{assoc}: \prod_{a, b, c, d: C} f: \text{Hom}_{C}(a, b) \times \text{Hom}_{C}(b, c) \times \text{Hom}_{C}(a, c) \)
6. \( \text{ident}: \prod_{a, b: C} f: \text{Hom}_{C}(a, b) \times (f \circ 1_{a} = f) \times (1_{b} \circ f = f) \)

We can now show that precategories are \( \infty \)-elementary in the sense that they are axiomatizable, up to equivalence, by \( \mathbb{T}_{\text{cat}} \) over the \( \infty \)-signature \( \mathcal{L}_{\text{cat}} \). In what follows below we will be making free use of the HoTT version of the Axiom of Unique Choice (AUC) ([171], Corollary 3.9.2).

\textbf{Proposition 5.7.1.} \textbf{PreCat} \simeq \textbf{Mod}(\mathbb{T}_{\text{cat}})

\textbf{Proof.} The proof boils down to proving, using AUC, that an axiomatization of a category in terms of a relation of composition is equivalent to the usual axiomatization
in terms of an operation of composition. First we define a function

\[ p: \text{Mod}(\mathbb{T}_{\text{cat}}) \to \text{PreCat} \]

So let \( C: \text{Mod}(\mathbb{T}_{\text{cat}}) \) and write \( t_i \) for the inhabitants of each axiom \( T_i \) that is part of the data of \( C \). We need to provide the data for conditions (1)-(6) in the definition of precategories. We are given \( O \) and \( A \) and those immediately take care of conditions (1) and (2). For condition (3) we first observe that \( \Sigma_{i: A(x,x)} I(i, x, x) \) is a mere proposition for any \( x: O \). For suppose that \( (i, \phi) \) and \( (j, \psi) \) are two terms of type \( \Sigma_{i: A(x,x)} I(i, x, x) \). To show that \( (i, \phi) = (j, \psi) \) it suffices to show that there is \( p: i = j \) and that \( p_*(\phi) = \psi \).

By applying \( t_5 \) to the data \( \langle x, i, j, \phi, \psi \rangle \) we get a proof that \( i = j \), i.e. a term \( p: i = j \).

Clearly, since \( I(i, x, x) \) and \( I(j, x, x) \) are mere propositions, we also get that \( \psi = p_*(\phi) \) and therefore we get that that \( (i, \phi) = (j, \psi) \) and therefore that the type \( \Sigma_{i: A(x,x)} I(i, x, x) \) is a mere proposition. By AUC and \( t_1 \) we get a term

\[ u: \prod_{x: O} \Sigma_{i: A(x,x)} I(i, x, x) \]

Thus we can define, for each \( x: O \), the following term

\[ 1_x =_{\text{def}} \text{pr}_1(u_x): A(x, x) \]

and thus we obtain a term

\[ 1^C =_{\text{def}} \lambda x. 1_x: \prod_{x: O} A(x, x) \]
as required by condition (3). Condition (4) follows similarly and we omit the details. For condition (5), let \( x, y, z, w: O \) and \( f: A(x, y), g: A(y, z) \) and \( h: A(z, w) \). Now let

\[
r = \text{def} \ (t_4)_{x,y,z,w,f,g,h} \circ (g \circ f), x, y, w)
\]

We can then define

\[
p_1 = \text{def} \ f_{2}(c_{x,y,z,f,g})
\]

\[
p_2 = \text{def} \ f_{2}(c_{x,z,w,g,h})
\]

\[
p_3 = \text{def} \ f_{2}(c_{y,z,w,g,h})
\]

and thus \( r_{p_1,p_2,p_3}: \circ (f, h \circ C g, h \circ C (g \circ C f), x, y, w) \). But by definition we have a term

\[
\pi: \circ (f, h \circ C g, (h \circ C g) \circ C f, x, y, w)
\]

and therefore we get a term

\[
(t_{11})_{x,y,w,f,h} \circ (g \circ f), \text{refl}_f, \text{refl}_h, \text{refl}_g, (h \circ C g) \circ C f, x, y, w, \pi: h \circ C (g \circ C f) = (h \circ C g) \circ C f
\]

Thus we can define

\[
\text{assoc}^C_{x,y,z,w,f,g,h} = \text{def} \ (t_4)_{x,y,w,f,h} \circ (g \circ f), \text{refl}_f, \text{refl}_h, \text{refl}_g, (h \circ C g) \circ C f, x, y, w, \pi
\]

and this gives us the section

\[
\text{assoc}^C: \Pi_{a,b,c,d: C_f: \text{Hom}_C(a,b)} \Pi_{g: \text{Hom}_C(b,c)} \Pi_{h: \text{Hom}_C(c,d)} \quad h \circ (g \circ f) = (h \circ g) \circ f
\]
as required by condition (5). Condition (6) follows similarly. So we can now write \( p(C) \) for the precategory given by the data

\[
(O, A, 1^C, \circ^C, \text{assoc}^C, \text{idl}^C, \text{idr}^C)
\]

with the notation as in the proof of Proposition 5.7.1.

Conversely, we need to define a function

\[
q: \text{PreCat} \to \text{Mod}(\mathbb{T}_{\text{cat}})
\]

So let \( C \) be a precategory given by the data

\[
(C, \text{Hom}, 1, \circ, \text{assoc}, \text{idl}, \text{idr})
\]

Given 1 we know that for each \( x: C \), \( \text{Hom}(x, x) \) is inhabited since \( 1_x: \text{Hom}(x, x) \). Thus we can define

\[
I^C = \text{def} \lambda x. (\lambda f. (f = 1_x)): \Pi_{x: C} \text{Hom}(x, x) \to \text{Prop}_{\text{ul}}
\]

where we know that \( f = 1_x \) is a mere proposition since \( \text{Hom}(x, x) \) is an \( h \)-set. Exactly analogously, since \( \circ \) ensures that each type \( \text{Hom}(x, z) \) will be inhabited given \( f: \text{Hom}(x, y) \) and \( g: \text{Hom}(y, z) \) we define

\[
\circ^C: \Pi_{x,y,z: C} \text{Hom}(x, y) \to \text{Hom}(y, z) \to \text{Hom}(x, z) \to \text{Prop}_{\text{ul}}
\]

The verification of the axioms for this data is then entirely straightforward and we omit the details. So we can now write \( q(C) \) for the \( \mathbb{T}_{\text{cat}} \)-model given by the data

\[
(C, \text{Hom}, I^C, \circ^C, t_1^C, t_2^C, t_3^C, t_4^C, t_5^C, t_6^C, t_7^C)
\]
It is now easy to check that $p$ and $q$ are quasi-inverses establishing the required equivalence.

A strict category ([171], Definition 9.6.1) is a precategory in which the type of objects is an $h$-set. We write $\text{StrCat}$ for the type of strict categories. Let $\mathcal{L}_{\text{strcat}}$ be the signature whose category part is the same as $\mathcal{L}_{\text{cat}}$ but with $d(O) = 1$. Let $T_{\text{strcat}}$ be the $\mathcal{L}_{\text{strcat}}$-theory that contains the same axioms as $T_{\text{cat}}$ together with the following axiom which expresses that $O$ is an $h$-set:

\[(8) \forall x, y: O. \exists p: x =_O y. \forall q: x =_O p =_O q\]

We can now show that strict categories are “1-elementary.”

**Corollary 5.7.2.** $\text{StrCat} \simeq \text{Mod}(T_{\text{strcat}})$

**Proof.** Given Proposition 5.7.1 it remains to check that the new axiom (8) for strict categories ensures that $O^M$ is an $h$-set for any $T_{\text{strcat}}$-model $M$. By Hedberg’s Theorem ([171], Theorem 7.2.1) it suffices to prove that

\[(*) \prod_{x: O^M} \prod_{p: \text{Id}_{O^M}(x,x)} p = \text{refl}_x\]

is inhabited. Now suppose we are given an inhabitant of the (untruncated) translation of axiom (8)

$$\eta: \prod_{x,y: O^M} \sum_{p: \text{Id}_{O^M}(x,y)} \prod_{q: \text{Id}_{O^M}(x,y)} p = q$$

Setting $\eta_{x,x} \equiv \langle c, \theta \rangle$ and noting that $\theta_{\text{refl}_x}$ is a proof that $c = \text{refl}_x$ we get an inhabitant of $(*)$. But by the definition of our semantics we know that $\text{Id}_{O^M}(x,x)$ is an $h$-set and therefore $(*)$ will be an $h$-prop. Therefore, the proper (truncated) translation of axiom (8) will also produce an inhabitant of $(*)$, and we are done. \[\square\]
More interestingly, we can show that univalent categories are also “1-elementary” although we have to expand our signature to do so. A univalent category ([171], Definition 9.1.6) is a precategory satisfying the following additional datum, which expresses the fact that the canonical map \( \text{idtoiso}_{a,b} : a = b \to a \cong b \) is an equivalence for all \( a, b \):

\[
(7) \quad \text{cat} : \Pi_{a,b \in C} \text{isequiv}(\text{idtoiso}_{a,b})
\]

We write \( \text{Unicat} \) for the type of univalent categories. Now let \( \mathcal{L}_{\text{ucat}} \) be the following 1-signature

\[\begin{array}{llllll}
-1 & o & I & =^1_A & U & =^2_O \\
0 & t_0 & i & u_1 & t_1 & u_2 \\
1 & d_0 & d_1 & s_1^O & r_1^O & s_2^O \\
\end{array}\]

subject to all the same relations as \( \mathcal{L}_{\text{cat}} \) as well as the additional relation \( t_1^O u_1 = s_1^O u_2 \).

We can then define \( \mathbb{T}_{\text{ucat}} \) as the \( \mathcal{L}_{\text{ucat}} \)-theory given by the axioms of \( \mathbb{T}_{\text{cat}} \) together with the following extra axioms:

\[
(8) \forall x, y : O. \forall f : A(x, y). \text{Iso}(f) \to (\exists! p : x =^1_O y. U(f, p))
\]

\[
(9) \forall x, y : O. \forall f : A(x, y). \forall p : x =^1_O y. U(f, p) \to \text{Iso}(f)
\]

\[
(10) \forall x : O. \forall f : A(x, x). \forall p, q : x =^1_O x. (I(f, x) \land U(f, p) \to r_1^O(p, x))
\]

\[
(11) \forall x, y : O. \forall f : A(x, y). \forall p, q : x =^1_O y. (U(f, p) \land U(f, q) \to p =^2_O q)
\]

where we have used the following abbreviations:

\[
\text{Iso}(f) \equiv \exists g : A(x, y) \exists h_1 : A(x, x) \exists h_2 : A(y, y).
\]

\[
\circ (f, g, h_1) \land \circ(g, f, h_2) \land I(h_1) \land I(h_2)
\]

\[
\exists p : x =^1_O y. U(f, p) \equiv \exists p : x =^1_O y. (U(f, p) \land (\forall q : x =^1_O y. (U(f, q) \to p =^2_O q)))
\]
Thus, axioms (8)-(10) express that $U$ is a bijective relation between isomorphisms and paths that sends identity to reflexivity and axiom (11) expresses that $U$ a functional relation. We now obtain the following.

**Proposition 5.7.3.** $\text{UniCat} \simeq \text{Mod}(\text{T}_{\text{ucat}})$

**Proof.** From Proposition 5.7.1 we can assume that the data for a model of $\text{T}_{\text{ucat}}$ is given by the same data as that of for a precategory, together with the interpretation of $U$. Given AUC, for any given $\text{T}_{\text{ucat}}$-model $\mathcal{M}$ we can extract from $U^\mathcal{M}$ a section

$$u^\mathcal{M} : \prod_{x,y : O^\mathcal{M}} \text{Id}_{O^\mathcal{M}}(x,y) \to \text{Iso}^\mathcal{M}(x,y)$$

where

$$\text{Iso}^\mathcal{M}(x,y) \equiv \prod_{f : A^\mathcal{M}(x,y)} \text{isiso}(f)$$

such that

$$\pi : \prod_{x,y : O^\mathcal{M}} \text{isequiv}(u_{x,y})$$

Thus we get

$$\text{Mod}(\text{T}_{\text{ucat}}) \simeq \sum_{C : \text{Precat}} \prod_{x,y : O^C} \text{isequiv}(u_{x,y})$$

We can now take $\text{Unicat}$ to be the type

$$\sum_{C : \text{Precat}} \prod_{x,y : O^C} \text{isequiv}(\text{idtoiso}_{x,y})$$

There is then a natural map $f$ from $\text{UniCat}$ to $\text{Mod}(\text{T}_{\text{ucat}})$ which sends $\langle D, \text{univ} \rangle$ to $\langle D, \text{idtoiso}, \text{univ} \rangle$. For a given

$$\langle D, u, \pi \rangle : \text{Mod}(\text{T}_{\text{ucat}})$$
the homotopy fiber of \( f \) over \( \langle D, u, \pi \rangle \) is given by

\[
\text{hfib}_f(\langle D, u, \pi \rangle) \equiv \sum_{(C, \text{univ})} \langle C, \text{idtoiso, univ} \rangle = \langle D, u, \pi \rangle
\]

To show that \( \text{hfib}_f(\langle D, u, \pi \rangle) \) is contractible it clearly suffices to show that for all \( x, y : O \), 
\( u_{x,y} = \text{idtoiso}_{x,y} \) and by function extensionality this reduces to giving an inhabitant of

\[
\prod_{x,y,p} u_{x,y}(p) = \text{idtoiso}_{x,y}(p)
\]

which by path induction reduces to providing an inhabitant of

\[
\prod_x u_{x,x}(\text{refl}_x) = \text{idtoiso}_{x,x}(\text{refl}_x)
\]

But by the axioms of univalent categories and of \( \mathbb{T}_{ucat} \) we get that both sides of the equation are (propositionally) equal to the (unique) identity map on \( x \). Thus \( f \) is an equivalence and we are done. \( \square \)

**Remark 5.7.4.** Clearly the axioms of \( \mathbb{T}_{ucat} \) employ full first-order logic. It is unclear whether \( \text{Unicat} \) can also be axiomatized in the coherent fragment over \( \mathcal{L}_{ucat} \).

### 5.8. Prospects

Let me conclude by saying a few words about the prospects of \( n \)-logic which are both technical and philosophical. On the technical side, Proposition 5.7.3 illustrates the kind of result that \( n \)-logic was designed to tackle, and which justifies the title of this Chapter. Namely, we want to tackle traditional model-theoretic questions (e.g. of axiomatizability) but about structures defined on homotopy types rather than sets. From this point of view, many future technical projects and questions suggest themselves. We list a few:
(1) Extending the syntax of $n$-logic to the case $n = \infty$ in such a way as to make it possible to define semantics as in Definition 5.3.1. (This is related to the well-known open problem of managing infinite chains of coherence data in HoTT.)

(2) Proving a general completeness theorem for all finite $n > 1$.

(3) Characterizing $n$-elementary types in general, i.e. proving a Loś Theorem for $n$-logic along the lines of [30], Theorem 4.1.12. (This is likely to be non-trivial even in the case of $n = 1$.)

On the other hand, $n$-logic opens up new vistas for investigations in formal philosophy. In particular, $n$-logic can be used as the formal justification for a novel theory of concepts. It thus allows us to take seriously the idea that a syntactically defined concept may have as its extension an “abstract shape” (i.e. a homotopy $n$-type). What could be the use of such a strange idea? Allow me to illustrate with one final example.

One of the founding problems of analytic philosophy (or, at least, of what we now call the “philosophy of language”) was how to make sense of the content of identity statements while maintaining the principle that equals could be substituted for equals. Frege’s seminal paper [45] attempts to solve this problem (roughly) by distinguishing between the sense of an expression and its reference. This distinction is a natural consequence of a very intuitive view on the nature of concepts – namely, that concepts have an extension that can be described in multiple ways (intensions). A formal theory of concepts that subscribes to this view inevitably leads to broadly set-theoretic systems (as indeed it led Frege to the Begriffsschrift). I think it is no exaggeration to say that such a view of concepts, given the success of set theory, has come to seem inevitable.

No more. With UF and $n$-logic we can re-imagine concepts as “abstract shapes”. This allows us to altogether do away with the distinction between sense and reference that had previously seemed inevitable. For in an abstract shape we may have distinct
points connected by paths. This can be seen as a formalization of the very basic idea that there are distinct intensions that are indistinguishable (with respect to the properties that hold of them). But crucially, we do so without assuming that there is some other third thing in addition to the intensions – their “reference” – that somehow makes it the case that they are indistinguishable.

To take a classic example, consider the following two concepts:

- FVM = “Heavenly body visible first in the morning”
- FVN = “Heavenly body visible first at night”

As is well known, FVM and FVN have the same reference. How to make sense of this? Simply take the extension of FVM and FVN to be an abstract shape that consists of two points and one or several paths between them. Indeed, we may call one of those points the “Morning star” and the other the “Evening star” and picture it as follows:

\[
\text{Morning Star} \quad \text{Evening Star}
\]

with the lines indicating paths. What could be the meaning of a path in such a set-up? We have spoken about interpreting the Morning Star (“MS”) and the Evening Star (“ES”) as points in an abstract shape, but not about what paths between them could mean. Well, firstly, what do paths do for us in the formal setting? A path $A \rightarrow B$ allows us to transfer any proof of a property that holds of $A$ to a proof that it also holds of $B$. In other words, it provides a method to transform a process for determining that some property holds of $A$ into a process for determining that that same property holds of $B$. Does this kind of thing translate to our informal case? I think it does, and quite
naturally too. What happens is that we determine that some property holds of the ES, say its position \( X \) relative to some other star \( S \) at a particular point in time. Now here is a method to transform this property of the ES into a property of the MS: use a telescope to pick out a distinguishing mark on the ES, and wait until morning. When the MS rises, use the telescope to pick out that same distinguishing mark. You can now confidently assert that the MS was also in position \( X \) relative to \( S \) at a particular point in time.

“But have you thereby demonstrated that the given property holds of the \textit{same} thing?”

The answer is: it is irrelevant! Do I \textit{know} that the ES is the same thing as the MS because I picked out the same distinguishing mark on both? No. What I do know (or, at least, what I am willing to defend) is that the property that I determined held for the ES (its position \( X \) relative to \( S \)) holds also of the thing that bears the same distinguishing mark I saw in the morning. This describes a process of transferring properties from the (intension) ES to the (intension) MS. Paths in abstracts shapes can thus be understood exactly as such processes. In other words, a path in this example is given by a process to transform a method that determines that the ES was in such-and-such position at this moment in time – by telescope and pencil, say – into a method for determining that the MS is in such-and-such position at this moment in time. What is the process? Carry out the telescope-and-pencil measurement, wait until morning, then carry out the same measurement on the bright spot that appears above the horizon. With bloodshot eyes, just as dawn is breaking, you will then be able to triumphantly declare: “This bright spot right there, \textit{at night}, appears at such-and-such position \( X \) relative to \( S \).”

Thus, after all the technical complexities, we have ended up with an extremely simple philosophical idea: the extension of a concept can profitably be understood as the abstract shape whose points are the concept’s various intensions and with paths indicating processes available to us for transferring properties of one intension to properties of
another. In such a view, there is no longer any distinction between sense and reference: each concept can instead simply be understood as a cluster of intensions (some of them connected by processes that allow us to transfer properties between them). We move, in other words, from a metaphysics of objects (or identity conditions on objects) to a metaphysics of processes/methods. What processes? Those that allow us to transfer properties of certain things to certain other things. These processes need not be unique, nor need they be definitive. Indeed there being multiple processes between things (or between other processes) is exactly what makes this picture powerful and interesting.

The above example of sense and reference is not meant as a complete account, nor is it put forward as a theory. I offer it merely as an illustration that simple philosophical ideas can now be backed up by a view of concepts as abstract shapes. It is in exactly this way, fortified by formal work, that the possibility of carrying out a new kind of “logical analysis” is established. If one wishes to remain bound to a “normal” formal philosophy – to be carried out in terms of research projects and subject to pre-existing grand narratives – it is imperative that we explore its implications. Indeed, up to this point in this dissertation (i.e. up to this very sentence), I submit that everyone should follow me – only some perverse parochialism would prevent philosophers from exploring category-theoretic thinking and the mathematical logic of UF as the basis for an alternative philosophical logic, to be profitably applied to all kinds of rigorous philosophical pursuits.

For me this is not enough. I wish to go further. I want to view categorical thinking and the mathematical logic of UF as the basis for an “abnormal” philosophy, i.e. as grounding a philosophical logic that can reshape the very form in which philosophy is written and communicated. But into this strange wilderness, I can no longer reason that everyone has to follow me. There is no rational reason to do so. I cannot argue for it. I can merely present the wilderness to you, and let you decide for yourself. So, to those
unafraid of being lost in the woods in the waning evening light I extend an invitation to
join me for one last long dance in the sheaf-shaped Appendix at the end.
Appendix in the Shape of a Sheaf

What follows is an essay on the method of Wittgenstein in his late period offered as an example of a philosophical outlook that can only be presented in a “categorical” rather than “deductive” style. Allow me to explain the structure of the essay and how it is meant to be read. The main text of the essay begins on p. 316 and ends on p. 335. Each page in this range is split into two halves: a passage at the top half of the page (whose beginning is indicated by “→”) and a passage at the bottom half of the page (whose beginning is indicated by “←”). The top half of each of pp. 316-335 can be read as a linear, “objective/impersonal” account of the method of the late Wittgenstein, moving from left to right. A proper abstract for this part of the essay would be something like this:

“I investigate the method of the late Wittgenstein. I begin by pointing out certain similarities and dissimilarities between his early and late work. His ideas of the mystical, I claim, remain the same but his conception of “nonsense” undergoes a radical change. I then argue that the distinction between description/explanation in his late work functions very similarly to the saying/showing distinction of the Tractatus. Framed in this manner, I claim that the late Wittgenstein wanted to explain (by describing) that philosophy is merely a kind of human behaviour. It is his inability to state this as a thesis (since he himself is engaging in this behaviour) that ultimately informs the method of his late work. I finish by considering some objections and by pointing out that there is a fruitful analogy between his method as I conceive of it and the use of a certain technical tool (“sheaves”) in contemporary algebraic geometry.”

On the other hand, the bottom half of pp. 316-335 is a “subjective/personal” account of Wittgenstein’s impact on me, but it reads from right to left (i.e. it starts on p. 335 and ends on p. 316). In addition, bottom and top are arranged so that the former can be read as a commentary on the latter. There are thus (at least) two “layers” of structure in this essay and therefore (at least) two (complementary) ways of reading it: firstly, by going left-to-right on top then right-to-left at the bottom; secondly, by going up-down in a zig-zag moving from left-to-right. The two diagrams on p. 315 and p. 336 merely express in diagrammatic form what I've just described. (The one on p. 315 indicates that each zig-zag is a “local section” that patches together into a “global section” indicated on p. 336.)

Finally, let me note that the peculiar structure of this essay is not meant as a mere exercise in “non-standard philosophy”. The form and the content of this essay were chosen together and are inseparable: this essay could no more be “linearized” as could the particular “non-linear” style in which it is written be about anything other than the late Wittgenstein and his impact on my thinking.

With respect to primary sources the following abbreviations are used: PI (Philosophical Investigations), RFM (Remarks on the Foundations of Mathematics), CV (Culture and Value), BLBK (Blue Book), BRBK (Brown Book), OC (On Certainty). As for the exclusion of other late works, the spirit with which I want to look at Wittgenstein’s late work is perfectly described by Stern:

[B]oth the Investigations and On Certainty, like Wittgenstein’s other posthumously published works, are much more accessible if one approaches them as
selections from a larger body of work. Looking at this larger body of work makes it easier to grasp the problems that occupied his attention. ([167] p. 446)
Descriptive Tools: Analogies/Language Games etc.

Perspicuous Representation

Natural Language → Object of Comparison

\[ p.316 \rightarrow p.317 \rightarrow \ldots \rightarrow p.335 \]
The question that concerns me is this: What was the method of the late Wittgenstein? To answer it, I begin with the notion of a sheaf from algebraic geometry. Suppose we have an object $X$ (or “shape” or “space”) that we want to investigate, e.g. a table. One way of carrying out such an investigation is to attach to every region in our space some particular batch of information, e.g. in the case of a table: hardness, discoloration, texture. Ideally, we want this rule to satisfy two conditions:

1) **Compatibility**: Information attached to overlapping regions of $X$ is the same. e.g./i.e. If I pick two regions $U,V$ on my table that overlap in some common region $W$ then we want the information attached to $W$ as a subregion of $U$ to be the same as the information attached to $W$ as a subregion of $V$.

2) **Local-to-Global (or the “Sheaf Condition”)**: For each way I choose to split up $X$ into regions $\{U_i\}_{i \in I}$ that cover it completely, there is a unique assignment of information to $X$ as a region of itself that is compatible with all the $U_i$. e.g./i.e. If I cover the whole table by some choice of its subregions then from the information attached to each of these subregions I can recover all the information attached to the whole table: we want the local information to be thorough enough such that only one unique table can be the object to which our rule applies.

A sheaf is thus a rule $F: \text{Space} \rightarrow \text{Structures}$ that is both compatible and satisfies a local-to-global condition and where a “space” is the object we are interested in studying and “structures” are those objects that we already “know enough about”. To construct a sheaf on some space allows us to “shift the discourse” from something less well-understood to something better-understood by ensuring that the less well-understood object is faithfully patched together from regions of better-understood structures – and this even though the better-understood structures in themselves may initially appear to have nothing to do with the less well-understood object that we want to study.

A sheaf can thus be regarded as an answer to the following question: what kind of rule turns local nonsense into global sense? More suggestively put: A sheaf is a rule to patch together nonsense into sense. And if we are in possession of a sheaf then we can literally replace our object of study with a sheaf defined on it – and proceed to study that sheaf, forgetting all about the object we started with. This simple but vague idea of a sheaf was of monumental significance in 20th century mathematics. Unrelated as it may seem, I must ask you to keep it in the back of your mind while we move to Wittgenstein. For what I want to claim, eventually, is that the method of the late Wittgenstein is exactly analogous to the method of studying spaces using sheaves.

---

There isn’t any particular relationship between the messages, except that the author has chosen them carefully, so that, when seen all at once, they produce an image of life that is beautiful and surprising and deep. There is no beginning, no middle, no end, no suspense, no moral, no causes, no effects [but only] the depths of many marvelous moments seen all at one time. (Kurt Vonnegut)
More precisely, what I aim to do is to show that the method of the late Wittgenstein is to construct a sheaf of objects-of-comparison defined on our natural language. But before any of that nonsense can be turned into sense, there are two aspects of Wittgenstein’s work that I want to draw attention to:

1. A continuity in the **purpose** of his late and early work.
2. A discontinuity in the **method** he used to achieve this purpose.

The early and late Wittgensteins are often treated as two distinct philosophers, originators of two irreconcilable schools of thought. This distinction has been put in many different ways: the early Wittgenstein was a “realist” while the later Wittgenstein was an “anti-realist”; a “logician” as opposed to a “mystic”; a “logical positivist” turned “radical critic” – and so on. There are undoubtedly many reasons to draw such distinctions, some of which even Wittgenstein would agree with, judging by his many explicit disavowals of the *Tractatus* in his later work (e.g. *PI* §114.) Yet we cannot ignore that there are also many important similarities between his late and early work. One such fundamental similarity is this: his conviction that philosophical propositions are a form of **nonsense** and that the aim of his philosophy is to unmask such philosophical propositions as nonsensical.

This conception of his “purpose” as a philosopher remained constant throughout his life. It was, we may say, the “fixed point of [his] real need” (*PI* §108) which was none other than complete clarity and the elimination of confusion. Compare for example:

> The right method in philosophy would be this. To say nothing except what can be said, [...], i.e. something that has nothing to do with philosophy. (*TLP* 6.54)

For the clarity that we are aiming at is indeed **complete** clarity. But this simply means that the philosophical problems should **completely** disappear. (*PI* §133)

In contrast with this continuity in his conception of the purpose of his work is a discontinuity in the **method** he saw fit for achieving it, i.e. for **revealing** philosophical nonsense as nonsense.

←. Which brings me, finally, to the real question: what really has been achieved here? What, really, have I described? To answer, consider the sentence above: “the method of the late Wittgenstein is to construct a sheaf of objects-of-comparison defined on our natural language”. It is a sinful sentence, if not outright nonsense and pretension bordering on idiocy.

The theory of sheaves and contemporary algebraic geometry offer no extra depth to this work. They are not meant to **substantiate** it in any way. Rather, this work **mimics** their methods. It is the *Harlequin* to their Macbeth. The same way analytic philosophy, once upon a time – before it became much too assured of its own method – used to mimic proof theory.

To think of Wittgenstein’s late philosophy as a sheaf is not meant to draw on the use of sheaves in algebraic geometry as a formal tool. It is not meant to **reveal** the technical structure of his thought. I am not using formalism **ideologically**, in order to suggest scientific substantiation. I am using it **stylistically**, **aesthetically**, even **playfully**, **comedically** and **provocatively** – and at the same time seriously.

I am envisioning here a “deep future” philosophy – that is all. Philosophy as a kind of science fiction, philosophy as a description of Humanity from the point of view of a joyful species that lies in its future. And all this, at the end of the day, merely in order to convalesce from Wittgenstein, from his RFM V §16. Which is no more than to say that the content and achievements of this sheaf-shaped essay is perfectly described by saying that – – – – – –
Someone who wishes to reveal philosophical nonsense as nonsense is immediately faced with the problem of how to do this without himself producing philosophical nonsense. How do you draw a circle around philosophical nonsense without adding to it? To reveal $X$ as $Y$ you need to regard $X$ from the “outside” – but this seems futile when it comes to philosophy.

Wittgenstein was well aware of the difficulty in the *Tractatus*: this is why he speaks of “throwing away the ladder” in *TLP* 6.54. He remains well aware of it even in his later period (though he is far less explicit about it):

I believe it might interest a philosopher, one who can think himself, to read my notes. For even if I have hit the mark only rarely, he would recognize what targets I had been ceaselessly aiming at. (*OC* §387)

His way of dealing with the problem is akin to a “skeptical solution”: he accepts the impossibility of the task, but thinks his method can bear fruit regardless. He accepts that he can reveal philosophical propositions as nonsensical only by producing nonsensical philosophical propositions – but he also believes that there is some kind of “mystical substratum”, some “inexpressible background against which whatever [we] could express has its meaning” (*CV*, p. 16) that vindicates his method. Wittgenstein saw his work as ultimately receiving its validation from some such background:

Is what I am doing really worth the effort? Yes, but only if a light shines on it from above. (*CV*, p.57) [My emphasis]

His conviction that he was himself producing nonsense that could only be redeemed by some “light from above” did not change between his early and late period. He did not come to regard the *Tractatus* as nonsensical in a way that his late work was not.

This, of course, raises a curious problem: in what sense could Wittgenstein have come to regard the *Tractatus* as flawed since he has already disowned the *Tractatus* as nonsense in *TLP* 6.54? The answer I want to give, roughly, is that the nonsense of the *Tractatus* was not positioned in the right way to receive the “light from above”. The *Tractatus*, as it were, did not consist of the “right kind” of nonsense (strange as this may sound). Whereas, presumably, his late work did? Indeed – but in order to make sense of these strange proclamations we need first to understand the shift in his view of what it means for a (philosophical) statement to be “nonsensical”.

So-called resolute and irresolute readings of the *Tractatus* are essentially competing attempts to grapple with this problem.

The point in all this of course is not to say: screw philosophy and let’s go save the poor. I don’t want to be guilty of such crassness, of such terrible bad taste.

Rather, what I am merely pointing out (even less: hinting at – like the dying invalid hints at the afterlife) in this sheaf of mine is a perspective (a *transspecial* one) from which poverty (respectively physical, chemical, psychological phenomena) makes sense – in the same way that kidney stones make sense (even before, let’s say, we figure out what causes them or how they can be alleviated).

“But all present-day economic theories (respectively physics, chemistry, psychology) do that too!” Yes but they do so by explaining. They do not do so in a “coordinate-free” manner in the same sense that e.g. general relativity makes sense of gravity in a “coordinate-free” manner – that is the difference. The fact that what I say may coincide with specific theories in specific co-ordinates (e.g. the present day) does not mean that they are competing theories. What I am saying is not a theory at all (I am no longer weaving cocoons).

The only thing we can be certain of, the only thing we can rely on, the only thing that the future will never look back on with scorn is description. And the only way mere description can ever prove sufficient (even worthy) is if a light shines on it from the deepest future.
What did “nonsense” mean for early and late Wittgenstein?

In the *Tractatus* nonsense was a failure to “give [...] meaning to certain signs in [our] propositions.” (*TLP* §6.53) Roughly: to say something nonsensical is to say something that, when (fully) analysed in some ideal notation, has failed to attach a meaning to one of its signs. For example, when we try to say something *about* saying – a trap he thought philosophers (including Russell and Frege) constantly fell into. (Consider, as an illustration, Wittgenstein’s criticisms of Russell’s type theory (*TLP* §3.331–3.334) on the grounds that Russell articulates rules for his notation rather than letting the notation speak for itself.) Nonsense, in short, was a violation of the rules of some ideal logical syntax underlying a failure to take account of the “logically atomistic” structure of the world.

For the late Wittgenstein, on the other hand, meaning essentially becomes identified with *use* (*PI* §43). Thus, a proposition is “nonsense” (meaningless) if it *fails* to have a use. Or, to put it in Wittgenstein’s own terminology: a proposition is nonsensical if it fails to “make a move” in a natural language game (if it fails, that is, to have any *use* in our everyday life). He once again saw philosophy as replete with such nonsensical propositions, which fail to be of any use, to make any contribution to any (natural) language game. This is why so many of his critiques in the *Investigations* end in variants of the phrase: “But that fails to say anything whatsoever”. For instance: “we are inclined to say something which gives no information” (*PI* §298), “[t]hat gets us no further” (*PI* §350), “[s]o it really [...] says nothing at all, but gives us a picture” (*PI* §352), “What are these words for? They serve no purpose” (*PI* §398), “that is not a significant proposition” (*PI* §408) etc.

These two differing conceptions of nonsense can help us shine a light on the corresponding methodological differences between late and early Wittgenstein. So we can ask: How could (early/late) Wittgenstein make manifest the fact that philosophical propositions are nonsensical (in the early/late sense)? I’ll take these questions in turn.

What constitutes a natural language game is no easy matter to settle - and this is of course a line of attack that can be levelled against the late Wittgenstein, who certainly seems to require such a distinction. For if nonsense is defined as a failure to make a move in any language game, then philosophical propositions surely don’t come out nonsensical: they have a use in exactly the language games in which they originate, namely in “philosophical” language games. (cf. *PI* §121)

What kind of fortress of ideas allows the ethicists to ignore urine in favour of argument maps? That kind of fortress is impregnable; it is impossible to raze it without inventing new weapons – without changing our very form of life. To overcome it – to even see the need to overcome it without scoffing the thought away – transpecial thinking is required. And once we see it – once we are struck by it – then we are no longer human.

Yes, this corridor was in some sense the birth of me – and the thought of it now has the texture of a field of snow at night, covered in shadows. As you walk past the shadows they stand still and look at you. Then you go back home and look out the window. The shadows have disappeared - but you remember their look – their look has finally struck you.

(As always, this is not a matter of understanding. It is a matter of understanding the seriousness.)
In the *Tractatus* the overcoming of nonsense is achieved by exhibiting the possibility of an ideal notation (one in which letters correspond to simples, facts are truth-tabular combinations of simples etc.). In this notation, Wittgenstein argues, no philosophical propositions could be coherently framed. And how do we come to see this fact? Roughly, when we grasp what this “ideal notation” would look like. Thus the overcoming of nonsense, through the *Tractatus*, is achieved by exhibiting (the possibility of) such an ideal logical notation which functions as a “yardstick of sense”: measured against this yardstick, philosophical propositions come out senseless.

A cautionary remark: the fact that this was how the Wittgenstein of the *Tractatus* understood nonsense does not mean that the *Tractatus* itself provided a general method with which any nonsensical proposition could be demonstrated to be nonsensical (in the above sense.) Cora Diamond puts this point nicely:

The fact that such remarks can be shown to be nonsensical does not mean that there is available any special *Tractatus* principle which would enable one to give a direct demonstration of the nonsensicality of such propositions [...] [T]he *Tractatus* does not provide a general principle which can be used to demonstrate that philosophical propositions are nonsensical [...] [but its] approach to philosophical confusion is [...] piecemeal; it depends essentially on enabling a person to see that the attempt to clarify the use of his or her words falters.([36], p. 206)

That is to say, Wittgenstein’s goal of unmasking philosophical propositions as nonsensical is achieved indirectly in the *Tractatus*. A philosophical proposition is proved nonsensical if it fails to be analyzed in the ideal logical notation; but not necessarily because some explicit violation of that syntax can be uncovered by a general method. For the *Tractatus* certainly does not develop such a general method.\(^a\)

\(^a\)For more on this point cf. [36], pp. 201-206.

There is a narrow corridor at Penn Station linking the subway with the NJ Transit platforms. The walls of this narrow corridor were always lined with vagrants and bums, wrapped in trash bags, sleeping on cartons, hugging their few belonging crammed into little trolleys, most of them soaked in their own urine, all of them half-crazy, mumbling to themselves, delirious. The great American workforce flowing in from New Jersey every morning would have to squeeze through this small tributary – this corridor of urine and bums. The sight hardly slowed anyone down. They passed it by, unflustered, like a river passes over the pebbles on its bed.

I had to pass through it too – going upstream, into New Jersey. Every day I had to go through this corridor of urine to end up in the Ivory Tower Room at Princeton to engage in a calligraphy of argument maps, a feast of cream-cheese bagels and grapes – all the while a million people pass through the urine-soaked bums of Penn Station (me included) without ever being struck. (Of course they understand.)

And I too would pass them by as I got urinated out into New York in the evening. But now I must pass them...as one passes kidney stones. These vagrants are the kidney stones that no system of ideas has managed to pass. And just as microscopic kidney stones are capable of incapacitating an entire organism, so these vagrants are capable of shutting down the kidney function of the whole of Humanity.

To see these pebbles as kidney stones: to be *struck* by a new “aspect”.

\(^\)
—→. So what about the late Wittgenstein? How can he make manifest the fact that a proposition “makes no move” in a language game? Simply put: by proving it useless. And how does he do that?

Firstly, he takes a philosophical proposition and regards it as a proposition that may be used in natural language, i.e. in our day-to-day life. Through a “perspicuous representation” (PI §122) the use of this proposition is regarded as analogous to the use of a homologous term in a “simplified” situation, namely a “language game” acting as an “object of comparison” (PI §130). This much is explicit in what he himself says of his work:

A perspicuous representation produces just that understanding which consists in ‘seeing connexions.’ [...] [It is] of fundamental importance for us. (PI §122)

[Language-games are set up as objects of comparison which are meant to throw light on the facts of our language [...] (PI §130)

But that is only half the task. We must now make our way from the “object of comparison” back to our natural language. For nothing we say about the object of comparison need necessarily apply to our natural language. Schematically:

\[
\text{Perspicuous Representation} \quad \overset{\text{Objects of Comparison}}{\longrightarrow} \quad \text{Natural Language}
\]

Notice that we have moved from the global (natural language) to the local (language games as objects of comparison). In order to get back, we need to “patch” together the local back into the global. In other words, we need to see the rule that takes us from our natural language to perspicuously represented objects of comparison as a sheaf.

And here lies the essential methodological difference between early and late Wittgenstein. Gone is the appeal to logic, so prevalent in the Tractatus. Logic is replaced by description, which now assumes a central role. The distinction between saying and showing in the early Wittgenstein is replaced by the distinction between description and explanation. It is on this latter distinction that the whole method of his late work is based. Description becomes the sayable and explanation becomes what is showable through the sayable. Just as what can be shown cannot be said (TLP 4.1212) what can be explained cannot be described – and just as what can be shown is revealed through the form of what can be said (logic) so is what can be explained revealed in the perspicuity of what can be described (language games as objects of comparison).

The early Wittgenstein’s method was to show by saying. The late Wittgenstein’s method is to explain in order to describe. But the late Wittgenstein can no longer rely on “logic” as the bridge between description and explanation. What can he rely on? This is the question we must answer if we are to figure out his method.

←−. How the inequality between the West and the rest of the world, for example, fails to strike anyone, even though everyone understands it.

And what is important here is that this kind of pronouncement must not be regarded as a proposition with a truth value, but it must be regarded much as Wittgenstein regarded his own propositions. As nonsense that can strike you.

Everyone can understand the above situation – anyone can throw explanations at it, clothe it in data and theories. But they do so at the cost of being struck by it – and the only appropriate reaction to it – once it’s stricken – is illness or suicide. Or the transpecial.

(I take here a sharp and only momentary turn to the “political” – because this is the right way to approach the political: as a sharp turn from the aesthetic, the philosophical, the literary. A “rational” approach to the political, on the other hand – we have a name for such a thing too: ideology. The political needs to be glimpsed, momentarily, as one glimpses a distant sea on a winding mountain road.)
So let us now get clear on what explanation and description meant for the late Wittgenstein. He repeatedly insists that what he is trying to do (what “we” must do) is to move from explanation to description. Indeed most of the methodological remarks he makes (about his own method) amount to restatements of that principle. He writes, for example, that “it can never be our job to reduce anything to anything, or to explain anything. Philosophy really is ‘purely descriptive’ ” (BLBK p. 18) and that “our method is purely descriptive; the descriptions we give are not hints of explanations” (BRBK p. 125) and, in a slightly different context that “what [he] has to do is something like describing the office of a king; in doing which [he] must never fall into the error of explaining the kingly dignity by the king’s usefulness, but [he] must leave neither his usefulness nor his dignity out of account” (RFM VII §3 (my emphasis)) and “one has to pass from explanation to mere description.” (OC §189) And finally, the most forceful statement of all, in the famous PI §109:

We must do away with all explanation and description alone must take its place. And this description gets its light, that is to say its purpose, from the philosophical problems. [...] The problems are solved, not by reporting new experience, but by arranging what we have always known. Philosophy is a battle against the bewitchment of our intelligence by means of our language.

We arrange what we already know (by describing it) – we do not postulate theoretical terms that could explain what we want to know. This is – in broad strokes – the fundamental methodological constraint of the late Wittgenstein: to do as much as possible using only descriptions. (And note, once again, the implicit local-to-global principle at play here: using only local descriptions we must somehow arrive at a “global” explanation.)

There is some ambiguity as to what the Wittgensteinian “we” is supposed to refer to in general. It can refer both to the generic confused philosopher that Wittgenstein is trying to cure (e.g. “What reason have we for calling ‘S’ the sign for a sensation?” PI §261 [first emphasis mine]), as well as to himself, in which case what “we” think and do is what he (Wittgenstein) thinks and does (e.g. “What we do is to bring words back from their metaphysical to their everyday use” PI §116) In all of his methodological remarks, like the ones above, the “we” can be unambiguously interpreted in the latter sense.

The undercurrent here is a radical expansion of the domain of what we take to be a “rigorous argument”. Initially, methods of argument in the new domain must be contrasted with methods in the old domain and must, in being so contrasted, necessarily appear as unrigorous, irrational. This is the point at which this essay is written – the contrast between this essay and a linear argument in analytic philosophy should be obvious. What is not yet obvious is that the method underlying this essay is also rigorous – but in an expanded sense, a sense that takes into account a new foundation and a new method in mathematics, of which sheaves in algebraic geometry are but one aspect.

“A logic of analogies, of metaphors, of mimicry!” Isn’t this really a kind of comparative literature clothed in robes of rigour, with category theory in the role of a false prophet? Yes and no: this is an expansion of rigour, not a rejection of what we now see as rigorous. The process is systematic: with mathematics, always, as a guide. (I am willing to say: without a new vision of mathematics, there can be no new vision of philosophy. Perhaps this means that my soul belongs to Plato rather than to Aristotle.)

(All this must surely sound “crazy” now. “Transpecial perspective”, “forget about evolution”, “mimicry as methodology” etc. And they will remain crazy unless I supplement them with a suitable expansion of a notion of “rigour” that can accommodate them. This I intend to do – more than that: to this I am devoted.)
The importance of description for the late Wittgenstein is that it ensures the perspicuity (PI §122) of the representations (language games) that he employs. Any such representation that involves explanation (theoretical propositions, generalizing principles etc.) fails to be perspicuous. Explanation adds into a representation unnecessary, ideal elements (e.g. mental states) – instead of focusing on what is already immediately available to us: “Since everything lies open to view there is nothing to explain.” (PI §126)

But Wittgenstein is not merely interested in the description of the use of words in language – he is no “ordinary language philosopher” even when he writes things like “[w]hat we do is to bring back words from their metaphysical to their everyday use.” (PI §116) For he clarifies:

Our method is not merely to enumerate actual usages of words, but rather deliberately to invent new ones, some of them because of their absurd appearance. (BLBK p. 28) [my emphasis]

Namely, Wittgenstein does not limit himself to descriptions of natural language as he finds it. He allows himself more tools: descriptions of analogies and dysanalogies; similarities and disimilarities; descriptions of imaginary as well as “realistic” language games. All of these, and more, are used by Wittgenstein. And they are all, if read correctly, descriptive and never explanatory.

What is important therefore is that only description takes place in his philosophy. Descriptions are not preludes to a general explanation. They are assembled there for the reader, who must then turn them into an explanation – but they do not offer an explanation. Only the reader can turn these descriptions into explanations that may enlighten him:

Anything your reader can do for himself leave to him. (CV p. 77)

I ought to be no more than a mirror, in which my reader can see his own thinking with all its deformities so that, helped in this way, he can put it right. (CV p. 18 (1931))

That is not to say that “bringing a word back to its everyday use” may under no circumstances consist solely in the thorough description of its everyday use. But this might not always be the case.

Discovery and science must become secondary to the enterprise of mimicry, in exactly the same way that philosophy became secondary to discovery and science at the dawn of the technological age.

I am asking you to imagine the age where scientific discoveries that would have in other times been considered epoch-making will greeted in much the same way that a new book on metaphysics is greeted now. For then limits will have been placed on physics in the same way that limits have now been placed on philosophy.

The notion of a “scientific discovery” will seem then as laughable as the notion of a “philosophical discovery” seems to us now. Imagine that kind of situation - can you? That is where transpecial philosophy begins.

But crucially: it is not that this outlook must be grasped, but rather it must strike.

A time when humanity cares only about description and not at all about explanation (a time in which thinking has truly become a kind of behaviour): this is the “deep future”. It is only against the light of this deep future that I was able to overcome RFM V §16. (It is also, I think, only against such a background that Wittgenstein’s late philosophy can be salvaged from skepticism or quietism.)
But is pure description really sufficient for Wittgenstein’s purposes? Can we really only rely on these “descriptive tools” to make our philosophical points? If pure description is what makes a representation perspicuous, then what is it about description that “produces just that understanding which consists in ‘seeing connexions’”? (PI §122)

Take PI §409 as an example:

Imagine several people standing in a ring, and me among them. One of us, sometimes this one, sometimes that, is connected to the poles of an electrical machine without our being able to see this. I observe the faces of the others and try to see which of us has just been electrified.

Wittgenstein is here analyzing the relation between the grammars of the words “know” and “pain”. Observe how he goes about doing that. He is not simply describing the use of “know” in everyday language; he also provides us with a description of an imaginary situation. His remark is meant to be taken not in the spirit of: “Let me explain to you the meaning of the word pain by imagining this scenario” but rather in the much more open-ended “Think of what comes to mind when you hear the word ‘pain’. Now let me describe a game for you.”

But why should I understand such descriptions of objects of comparison (fictional or otherwise) as related to the grammar of the word “know”? Perspicuous they may be, but how does that make them philosophically relevant? Why should I understand them as related to epistemic concerns about introspective knowledge?

I cannot: pure description is indeed not enough to “see connexions.” But Wittgenstein is aware of that: consider the second sentence in PI §109: “[T]his description gets its light, that is to say its purpose, from the philosophical problems.” Wittgenstein admits that it is not really pure description that will do the trick of getting us to the relevant object of comparison. It has to be pure description seen in the light of a philosophical problem. The problem and the description have to first exist side-by-side – and then one must be struck by it.

We arrive at the key notion: being struck. This is the extra datum: this is what can turn a representation into a sheaf. This is what can shed the “light from above”.

With the development of contemporary foundations of mathematics in the 20th century the dominance of philosophy over mathematics comes to an end (its dominance over physics perhaps ended even earlier). Free from philosophy, secure in its foundations, mathematics assumes a completely different form, in which even proof and rigour assume secondary importance, overshadowed by structure and construction. (Mathematics and philosophy have now entered a state of profound discord and misunderstanding. At root it is a methodological discord.)

This opens up room for a new act of mimicry: how to immitate new mathematics when doing philosophy. This – initially purely stylistic – attempt is what I have in mind when I speak of writing like a sheaf. The idea here is to immitate mathematics, but not to import its results in order to make points and prove theses. And this initially vacuous stylistic break needs a completely new background to substantiate it, to make it pack a punch: this is where the transpecial comes in.

In transpecial philosophy, we are no longer engaged in clarification. We are engaged in mimicry. (Mimicry and simulation is what a species must do after it develops a “theory of evolution”.)
To try and get a grip on what it means to “be struck”, let us take a typical example of a “descriptive proposition” that Wittgenstein employs in his own philosophy:

(P) Let us not forget this: when ‘I raise my arm’, my arm goes up. (PI §621)

Is (P) nonsensical? To answer this we must ask: does it have a use? Well, it very well could: If, for example, (P) is used to distinguish between someone with a phantom limb and someone with an actual limb. But what concerns us here is whether or not it has a use in the context in which Wittgenstein employs it. That is to say, in his philosophy. And here the answer seems to be no.

But if propositions like (P) do not have a use, then what is left? What can they possibly accomplish within the context of Wittgenstein’s philosophy? Does (P) cause behaviour? Does someone, e.g., raise their arm or fetch a slab because of it? Not quite - more accurately, it is the kind of proposition that stops a conversation in its tracks: (P) is uttered; it strikes everyone as obvious, as almost a joke; no-one reacts, except perhaps to ask for clarification. But the very point of something like (P), for Wittgenstein, is that we must allow it to strike us as remarkable: “Don’t look at it as a matter of course, but as a most remarkable thing.” (PI II.x)

(P) doesn’t have a use but it has an effect. This effect is achieved by seeing (P) in the light of a particular philosophical problem. The use of (P) does not consist in causing or responding to behaviour, engaging in our form of life, acting or reacting, etc. It consists in striking us, in the sense of PI §129: “[W]e fail to be struck by what, once seen, is most striking and most powerful.” “Being struck” by (P) requires no movement and no activity. It accomplishes nothing within our form of life (what it accomplishes transcends our form of life) and thus has no use within that part of it comprised of “late Wittgensteinian philosophizing”. It is nonsense – but nourishing nonsense.⁴

And we must now also come to see Wittgenstein’s whole late philosophy in the light of something – and allow it to strike us. This is the final piece of the puzzle – in what way is his late philosophy supposed to strike us?

⁴Of course, there is no fact of the matter as to what makes certain nonsense more nourishing than the rest. Attempting to state such a fact would be the late Wittgensteinian analogue of trying to state a proposition expressing the logical form of propositions. Logic is now replaced by descriptive propositions like (P) – and their capacity to strike can no more be described than logical form can be said.
—→. If descriptions are merely nourishing nonsense, then what is wrong with (philosophical) explanation? What right does Wittgenstein have to discard explanations i.e. the propositions of traditional “dogmatic” philosophy, his own included (e.g. “The world is all that is the case” (TLP 1), “The general form of a proposition is: \( (p, \xi, N(\xi)) \)” (TLP 6))? As mentioned on the left, what is it that makes the Tractatus nonsense the “wrong kind” of nonsense? Wittgenstein cannot simply resort to saying that explanations are nonsensical, since as we have seen his own descriptions come out just as nonsensical according to his late conception of nonsense. Surely his exhortations to abandon explanation in favor of description lose their force if it merely means abandoning one type of nonsense for another. Wittgenstein must require that there is something special about his “descriptive nonsense” – something that sets it apart from “explanatory nonsense”.

The answer is as simple as it is unsatisfying: what he has against explanation is that it is part of the sort of traditional philosophy he is reacting against, no less, no more. He cannot distinguish his own “descriptive” nonsense from the “explanatory” nonsense of traditional philosophy any more than he can distinguish “nourishing nonsense” from “mere nonsense” within his own work: for they both fail to have a use and it is by virtue of this fact alone that they become nonsensical. But what he can do is distinguish his philosophy from traditional philosophy, if merely in its spirit and style. (He cannot explain why his philosophy is any less nonsensical than the philosophy he wants to prove nonsensical – but he can describe how they differ in both spirit and style. And this allows us to place them side by side – and be struck by their difference.)

Thus, the late Wittgenstein turns to descriptions of our natural language in order to shift our thinking to a perspicuously represented object of comparison - perspicuous exactly because only description was involved in attaining it. Yet these descriptive tools also turn out to be nonsensical by Wittgenstein’s late standard. In order to use them to gain the insight aimed at, we must transcend them by allowing them to strike us.

Utter confusion! We have been led to a patchwork of objects of comparison – perspicuously represented – and yet we are now stuck. Nonsense here, nonsense there – what, really, is the point? What is it exactly that can shed “light from above” on this nonsense?

aIn an incomplete Preface to PI he begins by stating: “This book is written for those who are in sympathy with the spirit in which it is written. This is not, I believe, the spirit of the main current of European and American civilization.” (CV p. 6)

←−. And after I had come up with the symbol of the “future species”, after I had given the name “transpecial” to the “light from above” that could give my intellectual work its meaning once again – then I started finding evidence for it everywhere! I began reading voraciously again, and I was amazed to find that this exact same insight had been had by so many thinkers before me! This is what Kant thought the critical philosophy was to be applied to, this is what Nietzsche meant when he spoke about “the proposition ‘the species is everything, one is always none’” ([142], §1), this is what Kierkegaard meant when he spoke about this or that, this is what Badiou meant when he spoke about “generalized communism” etc. I found transspecial thinking everywhere around me! Hidden in plain sight!

Of course, I very soon realized that I had as much right to be amazed at all this as one has to be amazed to find the same reflection of oneself in different mirrors. For I was not finding my ideas in these texts – I was merely seeing my own ideas reflected on them.

Had I lost my capacity to read and learn? Perhaps – but at least I was no longer transparent. When before I saw nothing in the mirror that I held up against my ideas, I now saw something – except now everything had become a mirror of my ideas. Or, put more modestly: I has become a mirror myself – I was immitating everything around me.

And this, finally, is what led me to abandon the “evolutionary” model of knowledge in favour of a model in which “mimicry” becomes primary. It is the idea of “mimicry” that finally led me back to mathematics.
I want to regard man here as an animal; as a primitive being to which one grants instinct but not ratiocination. As a creature in a primitive state. Any logic good enough for a primitive means of communication needs no apology from us. Language did not emerge from some kind of ratiocination. (ÖC §475)

Wittgenstein scholarship tends to focus on the last sentence, which has acquired some well-deserved infamy. I should instead like to focus on the first (bolded) sentence. I think it is the clearest expression of the “transcendental” aspect of Wittgenstein’s late philosophy. It expresses exactly what I think Wittgenstein’s target was throughout his late writings: to make us regard ourselves as (mere) animals. And thereby to view our philosophizing as no more (or less) meaningful as the barking of dogs or the roaring of lions.

But crucially: not – as in the Tractatus – because “philosophy is beyond us” and we should therefore pass it by in silence, but because what we call philosophy is no more than a human form of roaring and barking. Wittgenstein is no longer urging us (as he did in the Tractatus) to abandon metaphysics because it goes beyond what can be said precisely (propositions of the empirical sciences) but he is urging us to let metaphysics strike us as a uniquely human form of barking or roaring. In other words, I urge we take his remarks on the “natural history” of human beings as seriously as possible: that there is simply nothing (not even philosophy) that human beings can do that is not merely just another kind of human behaviour.

An absolute triviality? Yes, but of the most serious kind! (The “thesis” is simple, even sophomoric. But to be struck by its seriousness? That is remarkably difficult. And that is the achievement of Wittgenstein’s late philosophy.)

Yet clearly we cannot view this as an explanation. We cannot view his philosophy as espousing this as a thesis. He can only deduce it, as it were, transcendentally.

“Transcendental”: a scary word - god knows I’ve both laughed and cried in its presence. Here we must confront it, in order to both demythologize it and recast it – in some sense to “naturalize” it but perhaps even better to biologize it.

The perspective we are aiming at is not the divine, unknowable, noumenal: it is the taxonomic, from which we view our species as one prior node in the evolutionary graph. But we have not jumped out of the graph, merely moved further along on it. The background against which I say we make sense of my “great unravelling” is the species-genus tree of our evolutionary history – and the “transcendental” aspect of it consists in looking ahead as well as behind, indeed it consists in thinking/philosophizing/speculating as if we could view homo sapiens in the same way that we can view homo neanderthalensis now.

This – in barest outline – is what I mean by the term transpecial. This is the perspective that I had to invent in order to escape from Wittgenstein. And in this transpecial perspective, the mystical, unknowable, noumenal background remains. Except it no longer has the form of some impenetrable darkness, but rather that of the impenetrable shadow of some unknown animal (the animal that belongs to the species that lies ahead in our “evolutionary” future).

(And let me say here that the introduction of the term “transpecial” is not made lightly. It hides behind it an immense amount of confusion caused in me when RFM V §16 knocked me off the Tower. It brings to my mouth the taste of blood and to my bones the ache of an old injury. It is not my own blood and it is not my own ache: I introduce this term because I am convinced that it points to a universal ache. But how exactly to share it? That is the question.)
What I see Wittgenstein as trying to do is, essentially, what was beautifully described by Cassirer as the attempt to consider man as an animal [...] which produces philosophies and poems in the same way as silkworms produce their cocoons or bees build their cells.¹

Such a perspective is nothing new. (Post-Darwinian thought often circles around this insight.) The problem this position faces is obvious and well-known: there is no “view from nowhere” from which it can be asserted. My contention, therefore, is that Wittgenstein’s late philosophy is an attempt to assert such a perspective, in order to explain (by describing) in what sense exactly philosophical propositions are nonsensical: they are the human equivalent of the threshing of a fly trapped in a fly bottle.

But if that’s all Wittgenstein is after – then why doesn’t he just state it? It’s easy enough: Cassirer did it. Yes, but Wittgenstein cannot: for in being stated as a thesis it immediately becomes contradictory (“the logocentric predicament”). Wittgenstein comes very close to stating it in many places. For example:

The evolution of the higher animals and of man, and the awakening of consciousness at a particular level. The picture is something like this: Though the ether is filled with vibrations the world is dark. But one day man opens his seeing eye, and there is light. [...] [The picture] already points to a particular use. This is how it takes us in. (PI II.vii)

In this difficult passage Wittgenstein appears to be saying that the picture itself is misleading: “it takes us in”. But in what way does it “take us in”? It sounds very much as if Wittgenstein is telling us that this picture is false and that therefore the “silkworm thesis” above is true. Convenient as it would be, we cannot take him to be doing any such thing. What we must see him as doing, instead, is applying his method to such statements themselves: for to advance the “silkworm thesis” as an explanation will lead us nowhere.

Just as the silkworm cannot “talk” about its cocoon through the process of weaving it, so Wittgenstein cannot assert the falsehood of this thesis through the human process of making assertions.

²[29], p. 20

With the Darwinian theory of evolution, with cosmology, with particle physics, with statistical science, Humanity has finally built for itself an ideological framework (a “conceptual scheme”) from which it can understand itself as the result of chance, of coincidence, of probability. The mistake we have made is to regard this insight as the beginning of the solution to the problem of the future of Humanity. In fact, this insight constitutes the very statement of the problem. For in order to understand Humanity as the product of chance does not mean we can see it as the product of chance. (Physics tells us that what we see is 4-dimensional curved spacetime. But physics can never tell us that we see it.)

We must still, somehow, create a direction for ourselves. The oft-repeated “worry” that in coming up with the theory of evolution we have thereby become immune to its influence on us says nothing more substantial than that a tribe which has come up with the notion of divine retribution has thereby become immune to being divinely retributed. Just as the notion of divine retribution spurred tribes to new means of organizing themselves and conceiving of their futures, so must modern evolutionary thinking spur us to new ways of conceiving of the future of Humanity. But this must be non-evolutionary thinking: it must be transspecial thinking.

And in transspecial thinking, mimicry is what provides direction – not the search for explanation, not progress, not evolution.
We are therefore faced with a methodological absurdity, similar to the one of the early Wittgenstein: we must somehow draw a circle around the human activity of “making assertions” without making an assertion. In the late Wittgenstein’s own terms, this reduces to “blowing up” objects of comparison to the size of our own natural language. But this is clearly a formal impossibility, an absurdity. But this is no mere “size issue”. This impossibility is not the impossibility of inhabiting a styrofoam model of a house as though it were real. For such a styrofoam model can be seen as a “first approximation” to a house, whereas an object of comparison is not even that, as Wittgenstein makes clear in PI §130. This impossibility, rather, is akin to that of trying to see in four dimensions. Or, perhaps, of seeing a 2-dimensional object as a 3-dimensional one.

What is required is that we leap out of his objects of comparison and return to our own life. But how? We know how to get to objects of comparison, namely through descriptive tools constitutive of perspicuous representations. But, once there, how do we get back?

The impulse, of course, is to try and explain. But explanation has already come to an end when we have placed ourselves inside an object of comparison. For within it, there can be no explanation about it. The language game is a brute fact: it is our form of life, it is “what has to be accepted – the given” (PI §192). So the only thing we are left with and which we must now try to do is to find a kind of wonder just by looking into the phenomena that are perspicuously represented by objects of comparison. As Malcolm puts it:

When we perceive the futility of trying to explain the phenomena, then we can focus on the phenomena themselves, and even be awakened to a kind of wonder at their existence. ([117], p. 73)

“But what does this wonder consist of? And where can it lead us to? And how do you awaken to it?” Aren’t all these questions merely seeking an explanation? And isn’t that that must be given up?

To awaken to a kind of “wonder”? You know what – in the silence that followed the passing of Alexandria, free from poetries and exultations, I found that it was possible to think, to do mathematics even. That mathematics itself is an activity just like walking did not mean I could not walk. And slowly, steadily, I did it: I started walking again.

But where to? I could not ask that: for I would be seeking an explanation (producing more cocoons). What, then, was left? Perhaps I could just think of myself as just one stage in the evolutionary development of mankind, no more no less. But isn’t that, too, an explanation? Am I not still weaving cocoons? Isn’t the “evolutionary” explanation just that – an explanation that pre-supposes an objective stance towards the whole of humanity, but is produced by it? And this kind of thinking quickly leads you back to an examination of different dogmas, a pursuit of explanations – thinking that Wittgenstein’s problem revealed to me as fruitless.

There were two options. Either the problem that had struck me had struck no-one else before or no-one else had ever thought of this as a problem. Neither seemed possible to me. Yet one thing was certain: the world around me still functioned – and the people around me still went about their business largely untroubled. Everyone around me seemed capable of walking. And so, as a beginning, I could start immitating them. If nothing else, I had a certain bedrock to lean on: a community of people around me that seemed to have either solved the problem or to have been left completely unmoved by it.

Mimicry, after all, provided a kind of direction – and one that required no explanation to justify it. (The child, for instance, mimics its parents not because it has found a reason to do so.) I could not ask whether or not anyone around me felt the same way about humanity, nor could I pursue a solution to my problem. But what I could do: I could mimic those around me who seemed to behave as though the problem had never occurred to them.
The proposed solution thus strikes us as almost a joke: we must wholly abandon the idea of seeking an explanation. What is there is simply this: the human form of life. And all we have (the given) is our own part in it. In other words, as a philosophy, we are entering the realm of the subjective.

This “subjective turn” is similar to what Kierkergaard urged, in a different context and with different problems in mind – and also to what Nietzsche had in mind when he wrote that [T]he great problems all demand great love [...] It makes the most material difference whether a thinker stands personally related to his problems, having his fate, his need, and even his highest happiness therein; or merely impersonally, that is to say, if he can only feel and grasp them with the tentacles of cold, prying thought. ([142], §345)

This is all similar, indeed, to all good irrationalists, but there it no lack of rationality here. This is exactly antiphilosophy in the sense of Badiou [13] (Kierkergaard, Nietzsche and Wittgenstein are all examples of what Badiou calls antiphilosophers.)

With fanfare and imagination, with an open mind and the taste of blood in our mouth, accompanied by an invisible procession – we enter the realm of personal experience. Wittgenstein is right there with us:

Working in philosophy - like work in architecture in many respects - is really more a working on oneself. On one’s own interpretation. On one’s way of seeing things. (CV p. 16 (1931))

We are entering the un-philosophical, the quasi-aesthetic, quasi-subjective, quasi-mystical, the ineffable, ridiculous, hymnal, the comical, the juvenile, the ethical: we seek now the “philosophical insight” that can only be had on the streets outside the cathedral and only after we have chucked the gospels in the bin.

So much the better! A sure sign that our head is not stuck in the sand! A sure sign that our heads are not stuck up a scientist’s backside!

The God Abandons Antony
When suddenly, at midnight, you hear an invisible procession going by with exquisite music, voices,
don’t mourn your luck that’s failing now,
work gone wrong, your plans
all proving deceptive – don’t mourn them uselessly.
As one long prepared, and graced with courage,
say goodbye to her, the Alexandria that is leaving.
Above all, don’t fool yourself, don’t say
it was a dream, your ears deceived you:
don’t degrade yourself with empty hopes like these.
As one long prepared, and graced with courage,
as is right for you who proved worthy of this kind of city,
go firmly to the window
and listen with deep emotion, but not
with the whining, the pleas of a coward;
listen – your final delectation – to the voices,
to the exquisite music of that strange procession,
and say goodbye to her, to the Alexandria you are losing.
Wait just a second! Let us rein in the excess for a moment. For we must pause to ask: even if one agrees with my “reading” of Wittgenstein why must we then immediately jump to Byronian romanticism? Even if this was the kind of thing that Wittgenstein really was after in his late philosophy, why can’t he be systematic about it? He certainly reads like someone who thought and wrote systematically. So why can’t there be a theoretical method that one can extract from his late philosophy that – if followed – would lead one to regard philosophy as the barking of dogs, the roaring of lions, the cocoon-weaving of silkworms etc.? For Wittgenstein certainly gives us many tantalizing clues as to what such a method might look like: “inventing new similes” (CV p.19), “philosophy as poetic composition” (CV p. 24), “to make it possible for us to get a clear view [...] [to put] everything before us” (PI §125,126).

But the key observation to make here is this: the (nonsensical) descriptive tools (analogies, similes etc.) which lead us to objects of comparison are at the same time the very source of our ailments! For recall some of Wittgenstein’s descriptions of the origins of our confusions:

[A] simile that has been absorbed into the forms of our language produces a false appearance. (PI §112)

A picture held us captive. And we could not get outside it, for it lay in our language and language seemed to repeat it to us inexorably. (PI §115)

But similes and analogies are exactly what Wittgenstein himself uses! As Baker summarizes the point: “[Wittgenstein’s method] could be called a kind of homeopathy: a way of ‘treating’ pictures with other pictures.” We fight language with language. Nonsense is revealed through nonsense. We attend to the workings of our language exactly in order to overcome its seductions (to fly out of the fly bottle) – the same way we protect ourselves from a virus by vaccinating ourselves with weakened forms of it. But this kind of “attending” is not the cure itself – it is only what triggers the immune response.

[C][15], p. 187

I had to strip these clownish exultations down and place them in front of a mirror and keep them there until they broke down in tears. I had to turn the mirror on myself not in order to understand, but in order to pity. I had to stare long and hard but not in order for a new “aspect” to reveal itself, but until what I saw could only fill me with grief and pity.

One of the strangest pronouncements I ever heard: “Knowledge begins with grief.” This is now how I do justice to these words, which carry the unmistakeable form of a penetrating insight.

I had to sit in front of a mirror long enough to be able to perceive my bliss (my exultations) as something worth grieving over, indeed as something devastating. There was no sudden insight, no “realization in a flash”, no discontinuity. There was rather an erosion, over a very long period of time, at no point of which was I aware of the changing shape of my reflection. But erode I did: and then – suddenly – I had become perfectly smooth, a perfect reflective surface able to bounce back any idea: a thermodynamic “white body” – but not a sceptic.

Looking in the mirror, I saw nothing: no idea could affect me but there was also no idea I could hold on to.

What became clear then: the proper philosophical insight is that which you do justice to, not attempt to communicate.
And we must also keep in mind that Wittgenstein’s targets were not philosophical theories. For to negate a theory is also to assert a theory: the negation of a philosophical thesis is itself a philosophical thesis. But for the late Wittgenstein, one ought never “advance theses in philosophy” (PI §128). And we can only see Wittgenstein as an anti-theorist by seeing him as a theorist, i.e. by carving out a position for him comprised of the negations of what he argued against and thus read a thesis into his work. This mistake is common: Crispin Wright and Michael Dummett reading him as an anti-realist, for example, or Kripke reading him as a sophisticated skeptic.

To illustrate, regarding Wittgenstein’s discussion of pain in the *Investigations* Kripke writes: “Clearly much more needs to be said here: a few sketchy and allusive remarks on the analogy between ‘I am in pain’ and a groan hardly give a complete theory...” That it doesn’t give a complete theory is true – but what Wittgenstein is asking us to do is to allow it to strike us as a complete theory. Imagine that all philosophy of mind is simply a kind of musicology of different types of groaning – this is what Wittgenstein is asking us to do. And from such a perspective, what does our philosophy look like? About as useful as a musicology of groaning.

The right way to read him is not as an anti-theorist but as an “anti-theory-ist”. His target is not this or that token philosophical theory but Theory and Theorizing in philosophy in general. His targets are the many ways in which philosophers theorize, generalize and explain instead of describing, looking and letting themselves be struck. And to make that point through philosophical language (a theoretical anti-theory), one must become a clown and a malcontent: for one is forced to communicate one’s mockery through its very target. This is in some sense philosophy as a kind of self-loathing: a “philosophical anti-philosophy.” But whatever it is – it is not theory-building. It is not problem-solving. It is a kind of clowning around – but of the most serious kind.

*a*[84], p. 145

*b*[13], p. 69

When a clown said to me: *To whom he was not announced, they shall see; and those who have not heard shall understand.* A clown made me understand without hearing, and this noise I pass on.

To understand me, you must close your ears to what I am saying.

For when you close your ears to him and focus only on his face, even the clown becomes a tragedian of the highest competence.

Let the clown explain – and you will only hear his jokes. But let the clown describe – and you may see his sadness.
Even so, the question remains: can we perhaps extract a single, unified “formula” (rather than “theory”) from Wittgenstein’s late philosophy? Wittgenstein of course doesn’t think so:

There is not a philosophical method, though there are indeed methods, like different therapies. (PI §133)

But can a single consistent practice can be extracted from his late philosophy? No, it cannot – for we can easily see that Wittgenstein’s actual philosophical practice cannot cohere with his method as he describes it. As Kenny accurately puts it: “[T]here were more things in his philosophy than could be confined within his metaphilosophy.”

Consider the following examples:

The meaning of a word is its use in language. (PI §43)

A philosophical problem has the form: “I don’t know my way about” (PI §123)

It is what human beings say that is true and false; and they agree in the language they use. (PI §241)

Essence is expressed by grammar. (PI §371)

An inference is a transition to an assertion; and so also to the behaviour that corresponds to the assertion. (PI §486)

The point here is this: Even if the only propositions Wittgenstein deemed worthy of philosophy were those that “everyone would agree to” (PI §128) he still could not write a philosophy that makes this point by using only such propositions. (The same way that, say, one cannot write a textbook on logic without employing some natural language to make it understandable.)

[79], p. 182

RFM V §16 had infected me – and I found myself bedridden. Then – there is no other way to put it – began really what was a grand and glorious unravelling: Once you have given up the ghost, everything follows with dead certainty, even in the midst of chaos. (Henry Miller.)

If you see even mathematics and logic as an activity, then you must be willing to face the (not so exquisite) music: and this involves giving up (or suspending) all wonderment, all thirst for understanding. There are tremendous consequences. You are left bruised, bloodied, intellectually comatose – and what’s more: foolish.

The problem, to be very precise, is this: if you come to regard the totality of your thinking life as a human activity, then what can your thinking life possibly accomplish over and above human activities? I saw myself as a silkworm which tries to think and speak but only manages to weave cocoons – and if the silkworm were to see itself as a silkworm we would still only see it as producing cocoons. And if the silkworm tried to avoid this problem, it would still be weaving cocoons. (Comparing ourselves to silkworms leaves us no better: we are merely weaving more cocoons.)

I had not, of course, uncovered a kind of universal problem, a “new form of skepticism”. The problem was not a dialectical one: it was a personal one. I was in possession of no conceptual scheme that could soften the blow: the web had been broken, and I fell through it and ended up destroyed. Yet somehow, I had to recover from it.

This is my “exegesis” of Wittgenstein: a description of the problem he made me feel – and my attempt to recover from it.
Wittgenstein was aware of this inconsistency. Often in his late work, when he traps himself into a theoretical-sounding aphorism, he immediately reprimands himself:

Isn’t what I am saying: any empirical proposition can be transformed into a postulate-and then become a norm of description. But I am suspicious even of this. The sentence is too general. One almost wants to say “any empirical proposition can, theoretically, be transformed...”, but what does “theoretically” mean here? It sounds all too reminiscent of the Tractatus. 

(OC §321)

So the prospects for drawing out a single consistent philosophical practice from his late work looks grim, partly because his own philosophical practice (what he wrote) does not fully cohere with what he seemed to suggest philosophical practice ought to be like. (Kant, for example, would be the paradigm of a philosopher who tried to make his own philosophy cohere with his views about what philosophy should do.) We are still, it seems, stuck inside objects of comparison – and we have no way of jumping back to real life. There is no “Doctrine of Method” that Wittgenstein can provide us with that makes it clear to us exactly how to transfer our insights about objects of comparison into insights about our natural language. (That would lead to a new kind of critical philosophy.)

What we do have is the exquisite music – the idea that he wants to make us regard ourselves as animals, as primitive people in a primitive state...But is that enough to salvage Wittgenstein’s philosophy? Exquisite though the thought might be, it is no more substantial than a musical phrase, heard in passing, from some “invisible procession” on the street below. (It is no more an idea than a poem inserted in the middle of a stream of thought is an argument.)

“What’s all this nonsense about music? We want to know: what is the point?”

The point is this: the method of the late Wittgenstein requires an altogether new metaphor. Visual metaphors have run their course. (We are near the end of the era of the mirror and the glass.) I don’t yet have the words to describe the metaphor – but what I do have is sheaves.

The very idea of a foundation of mathematics becomes superfluous – an extravagance! If mathematics is seen as an activity there is no more need to set foundations for it as there is, say, for the activity of walking, fetching and so on. In that vein consider also the following passage, in which Wittgenstein urges an analogous shift of perspective:

When I say: “If you follow the rule, this must come out,” that doesn’t mean: it must, because it always has. Rather, that it comes out is one of my foundations. (RFM VI §46)

And this made me see the doing of mathematics by humans as no more mysterious than what a cat does when she hesitates before leaping from one rooftop to the next. Wittgenstein, in short, wants us to see ourselves as cats, even when we’re scrawling formulas on blackboards, even when we calculate solutions to the field equations. But this, of course, cannot be argued for in the usual sense (what I’m describing is not the adoption of an “anti-realistic” view) – it cannot be communicated except perhaps as a joke, as a kind of prank. And that’s how I saw it too, initially: “Wittgenstein wants me to see myself as a cat!” But the joke, eventually, wears off. And what is left?

Thus it was for me: questions of “certainty” concerning mathematical propositions lost all their depth. The problem of the “nature” of mathematics had disappeared. But I was not “cured” – in fact this was the beginning of a disease (skepticism). For once you can get rid even of mathematics – once you can see even mathematics and logic itself as an activity – then what is left? If Wittgenstein was really aiming at a therapy, then he must’ve failed – for I felt more diseased than ever!
Recall the sinful and mysterious statement: Wittgenstein’s late philosophy is a sheaf.
We are now finally ready to do justice to it, right at the moment of maximum confusion. (Indeed, of maximum exhaustion.)

Let us think: What has Wittgenstein provided us with? A rule by which to attach to each part of our natural language certain information (our natural language is the “object” or “space” to which this sheaf is applied). This information consists in a patchwork of perspicuous representations in simplified, primitive settings (language games). We have, therefore, a rule taking us from the global to the local – and we must now regard it as a sheaf. This is the feat the late Wittgenstein requires of us. This is the act that gives his whole late philosophy its meaning (thereby rendering it trivial): the patchwork of representations he has provided form a sheaf: they collate local nonsense to global meaning.

Thus the method of the late Wittgenstein is this: to provide us with a constant flow of “local” representations as opposed to one final comparison of natural language against some impregnable standard. Think of how much the late Wittgenstein wrote: if you can no longer explain, then all you can do is describe more and more. (We move, in short, from the analytic ideal of a “perfect deduction” to the category-theoretic idea of representation.)

But that the method is adequate, that the rule of perspicuous representations is a sheaf: this is the mystical, the inexpressible. To see that our language is a patchwork of language games and no more is to see ourselves as silkworms or lions. And that cannot be argued for: explanation, at this point, comes to an end.

“But this,” you cry, “is so utterly juvenile! This is quietism not even worthy of a sophomore!” To this we can only respond as someone else did – someone occupied with the same kind of move from the objective to the subjective that we are about to undertake:

It is exactly [the] seeming triviality [of the task] that makes [it] infinitely hard, since the task itself does not beckon directly, in a way that promises support to the aspirant, and because the task works against him, so that it needs an infinite effort just to discover the task [...] ([80], p. 108)

Yes, indeed. We are talking about a leap of faith here (this is our sheafification), though our destination has nothing to do with God, and is still very much about humanity. Down we slide! But neither to Dadaism, nor to irrationalism, not to the ineffable or transcendental or the mystical, not to the continental or the analytic, not to postmodernism or to art – we slide, instead, to the transpecial.

←−. What kind of insight is encoded by this “sheafification” of Wittgenstein’s objects of comparison? Let me give you a remark that had this effect on me (what exquisite music it made me hear! And all of this is but the crudest attempt to describe it):

Strangely, it can be said that there is so to speak a solid core to all these glistening concept-formations. And I should like to say that what is what makes them into mathematical productions. (RFM V §16)

I took Wittgenstein to be making the following “trivial but infinitely hard” point: It is not that our solid core is mathematical, but that we call our “solid core” (whatever it may be) mathematics. We call mathematical exactly that which does justice, so to speak, to how we use the word “certainty” in our everyday life. It is not that we had an idea of what it is for something to be certain, and then figured out that mathematical propositions are of that sort. Mathematical propositions, it suddenly occurred to me, must not be seen as examples of certainty, but as part of the definition of “certainty”.

RFM V §16 has struck me – but not as an “argument” that establishes a “thesis”. It had struck me in the most personal way possible – and I found myself floored.

335
Descriptive Tools: Analogies/Language Games etc.

Natural Language \[\rightarrow\] Perspicuous Representation \[\rightarrow\] Object of Comparison

Transcendental Philosophy

\[p.316 \rightarrow p.335\]

Life \[\rightarrow\] Transspecial Philosophy

Experience
Bibliography


337
338


[163] M. Shulman, *Homotopy type theory should eat itself (but so far, it’s too big to swallow)*: Homotopy type theory blog post, 2014.


