

Roadmap

Warrick Macmillan

February 14, 2021

1 Introduction

The central concern of this thesis is the syntax of mathematics, programming languages, and their respective mutual influence, as conceived and practiced by mathematicians and computer scientists. From one vantage point, the role of syntax in mathematics may be regarded as a 2nd order concern, a topic for discussion during a Fika, an artifact of ad hoc development by the working mathematician whose real goals are producing genuine mathematical knowledge. For the programmers and computer scientists, syntax may be regarded as a matter of taste, with friendly debates recurring regarding the use of semicolons, brackets, and white space. Yet, when viewed through the lens of the propositions-as-types paradigm, these discussions intersect in new and interesting ways. When one introduces a third paradigm through which to analyze the use of syntax in mathematics and programming, namely Linguistics, I propose what some may regard as superficial detail, indeed becomes a central paradigm, with many interesting and important questions.

To get a feel for this syntactic paradigm, let us look at a basic mathematical example: that of a group homomorphism, as expressed in a variety of sources.

Definition 1 *In mathematics, given two groups, $(G, *)$ and (H, \cdot) , a group homomorphism from $(G, *)$ to (H, \cdot) is a function $h : G \rightarrow H$ such that for all u and v in G it holds that*

$$h(u * v) = h(u) \cdot h(v)$$

Definition 2 *Let $G = (G, \cdot)$ and $G' = (G', *)$ be groups, and let $\phi : G \rightarrow G'$ be a map between them. We call ϕ a **homomorphism** if for every pair of elements $g, h \in G$, we have*

$$\phi(g * h) = \phi(g) \cdot \phi(h)$$

Definition 3 *Let G, H , be groups. A map $\phi : G \rightarrow H$ is called a group homomorphism if*

$$\phi(xy) = \phi(x)\phi(y) \text{ for all } x, y \in G$$

(Note that xy on the left is formed using the group operation in G , whilst the product $\phi(x)\phi(y)$ is formed using the group operation in H .)

Definition 4 *Classically, a group is a monoid in which every element has an inverse (necessarily unique).*

We inquire the reader to pay attention to nuance and difference in presentation that is normally ignored or taken for granted by the fluent mathematician.

If one want to distill the meaning of each of these presentations, there is a significant amount of subliminal interpretation happening very much analagous to our innate linguistic usage. The inverse and identity are discarded, even though they are necessary data when

that Martin-Löf Type Theory, the constructive foundational project which underlies these proof assistants, is rejected by those who dogmatically accept the law of the excluded middle and ZFC as the word of God.

It should be noted, the constructivist rejects neither the law of the excluded middle nor ZFC. She merely observes them, and admits their handiness in certain situations. Excluded middle is indeed a helpful tool, as many mathematicians may attest. The contention is that it should be avoided whenever possible - proofs which don't rely on it, or it's corollary of proof by contradiction, are much more amenable to formalization in systems with decidable type checking. And ZFC, while serving the mathematicians of the early 20th century, is lacking when it comes to the higher dimensional structure of n-categories and infinity groupoids.

What these theorem provers give the mathematician is confidence that her work is correct, and even more importantly, that the work which she takes for granted and references in her work is also correct. The task before us is then one of religious conversion. And one doesn't undertake a conversion by simply by preaching. Foundational details aside, this thesis is meant to provide a blueprint for the syntactic reformation that must take place.

It doesn't ask the mathematician to relinquish the beautiful language she has come to love in expressing her ideas. Rather, it asks her to make a compromise for the time being, and use a Controlled Natural Language (CNL) to develop her work. In exchange she'll get the confidence that Agda provides. Not only that, she'll be able to search through a library, to see who else has possibly already postulated and proved her conjecture. A version of this grandiose vision is explored in The Formal Abstracts Project.

2 Preliminaries

[Todo : Give overview of tools and theory in the making of this thesis]

2.1 Grammatical Framework

...

2.2 Martin-Löf Type Theory

...

2.3 Agda

...

2.4 Natural Language and Mathematics

...

3 HoTT Proofs

3.1 Why HoTT for natural language?

We note that all natural language definitions, theorems, and proofs are copied here verbatim from the HoTT book. This decision is admittedly arbitrary, but does have some benefits. We list some here :

- As the HoTT book was a collaborative effort, it mixes the language of many individuals and editors, and can be seen as more “linguistically neutral”
- By its very nature HoTT is interdisciplinary, conceived and constructed by mathematicians, logicians, and computer scientists. It therefore is meant to interface with all these disciplines, and much of the book was indeed formalized before it was written
- It has become canonical reference in the field, and therefore benefits from wide familiarity
- It is open source, with publically available Latex files free for modification and distribution

The genesis of higher type theory is a somewhat elementary observation : that the identity type, parameterized by an arbitrary type A and indexed by elements of A , can actually be built iteratively from previous identities. That is, A may actually already be an identity defined over another type A' , namely $A \equiv x =_{A'} y$ where $x, y : A'$. The key idea is that this iterating identities admits a homotopical interpretation :

- Types are topological spaces
- Terms are points in these space
- Equality types $x =_A y$ are paths in A with endpoints x and y in A
- Iterated equality types are paths between paths, or continuous path deformations in some higher path space. This is, intuitively, what mathematicians call a homotopy.

To be explicit, given a type A , we can form the homotopy $p =_{x=_A y} q$ with endpoints p and q inhabiting the path space $x =_A y$.

Let’s start out by examining the inductive definition of the identity type. We present this definition as it appears in section 1.12 of the HoTT book.

Definition 5 *The formation rule says that given a type $A : \mathcal{U}$ and two elements $a, b : A$, we can form the type $(a =_A b) : \mathcal{U}$ in the same universe. The basic way to construct an element of $a = b$ is to know that a and b are the same. Thus, the introduction rule is a dependent function*

$$\text{refl} : \prod_{a:A} (a =_A a)$$

called **reflexivity**, which says that every element of A is equal to itself (in a specified way). We regard refl_a as being the constant path at the point a .

We recapitulate this definition in Agda, and treat :

```
data _≡'_ {A : Set} : (a b : A) → Set where
  r : (a : A) → a ≡' a
```

3.2 An introduction to equality

There is already some tension brewing : most mathematicians have an intuition for equality, that of an identification between two pieces of information which intuitively must be the same thing, i.e. $2 + 2 = 4$. They might ask, what does it mean to “construct an element of $a = b$ ”? For the mathematician use to thinking in terms of sets $\{a = b \mid a, b \in \mathbb{N}\}$ isn’t a well-defined notion. Due to its use of the axiom of extensionality, the set theoretic notion of equality is, no surprise, extensional. This means that sets are identified when they have the same elements, and equality is therefore external to the notion of set. To inhabit a type means to provide

evidence for that inhabitation. The reflexivity constructor is therefore a means of providing evidence of an equality. This evidence approach is distinctly constructive, and a big reason why classical and constructive mathematics, especially when treated in an intuitionistic type theory suitable for a programming language implementation, are such different beasts.

In Martin-Löf Type Theory, there are two fundamental notions of equality, propositional and definitional. While propositional equality is inductively defined (as above) as a type which may have possibly more than one inhabitant, definitional equality, denoted $- \equiv -$ and perhaps more aptly named computational equality, is familiarly what most people think of as equality. Namely, two terms which compute to the same canonical form are computationally equal. In intensional type theory, propositional equality is a weaker notion than computational equality : all propositionally equal terms are computationally equal. However, computational equality does not imply propositional equality - if it does, then one enters into the space of extensional type theory.

Prior to the homotopical interpretation of identity types, debates about extensional and intensional type theories centred around two features or bugs : extensional type theory sacrificed decidable type checking, while intensional type theories required extra bureaucracy when dealing with equality in proofs. One approach in intensional type theories treated types as setoids, therefore leading to so-called “Setoid Hell”. These debates reflected Martin-Löf’s flip-flopping on the issue. His seminal 1979 Constructive Mathematics and Computer Programming, which took an extensional view, was soon betrayed by lectures he gave soon thereafter in Padova in 1980. Martin-Löf was a born again intensional type theorist. These Padova lectures were later published in the “Bibliopolis Book”, and went on to inspire the European (and Gothenburg in particular) approach to implementing proof assistants, whereas the extensionalists were primarily emanating from Robert Constable’s group at Cornell.

This tension has now been at least partially resolved, or at the very least clarified, by an insight Voevodsky was apparently most proud of : the introduction of h-levels. We’ll delegate these details for a later section, it is mentioned here to indicate that extensional type theory was really “set theory” in disguise, in that it collapses the higher path structure of identity types. The work over the past 10 years has elucidated the intensional and extensional positions. HoTT, by allowing higher paths, is unashamedly intentional, and admits a collapse into the extensional universe if so desired. We now examine the structure induced by this propositional equality.

3.3 All about Identity

We start with a slight reformulation of the identity type, where the element determining the equality is treated as a parameter rather than an index. This is a matter of convenience more than taste, as it delegates work for Agda’s typechecker that the programmer may find a distraction. The reflexivity terms can generally have their endpoints inferred, and therefore cuts down on the bureaucracy which often obscures code.

```
data _≡_ {A : Set} (a : A) : A → Set where
  r : a ≡ a

infix 20 _≡_
```

It is of particular concern in this thesis, because it highlights a fundamental difference between the linguistic and the formal approach to proof presentation. While the mathematician can whimsically choose to include the reflexivity argument or ignore it if she believes it can be inferred, the programmer can’t afford such a laxidasical attitude. Once the type has

been defined, the argument structure is fixed, all future references to the definition carefully adhere to its specification. The advantage that the programmer does gain however, that of Agda’s powerful inferential abilities, allows for the insides to be seen via interaction window.

Perhaps not of much interest up front, this is incredibly important detail which the mathematician never has to deal with explicitly, but can easily make type and term translation infeasible due to the fast and loose nature of the mathematician’s writing. Conversely, it may make Natural Language Generation (NLG) incredibly clunky, adhering to strict rules when created sentences out of programs.

[ToDo, give a GF example]

A prime source of beauty in constructive mathematics arises from Gentzen’s recognition of a natural duality in the rules for introducing and using logical connectives. The mutually coherence between introduction and elimination rules form the basis of what has since been labeled harmony in a deductive system. This harmony isn’t just an artifact of beauty, it forms the basis for cuts in proof normalization, and correspondingly, evaluation of terms in a programming language.

The idea is simple, each new connective, or type former, needs a means of constructing its terms from its constituent parts, yielding introduction rules. This however, isn’t enough - we need a way of dissecting and using the terms we construct. This yields an elimination rule which can be uniquely derived from an inductively defined type. These elimination forms yield induction principles, or a general notion of proof by induction, when given an interpretation in mathematics. In the non-dependent case, this is known as a recursion principle, and corresponds to recursion known by programmers far and wide. The proof by induction over natural numbers familiar to mathematicians is just one special case of this induction principle at work—the power of induction has been recognized and brought to the fore by computer scientists.

We now elaborate the most important induction principle in HoTT, namely, the induction of an identity type.

Definition 6 (Version 1) *Moreover, one of the amazing things about homotopy type theory is that all of the basic constructions and axioms—all of the higher groupoid structure—arises automatically from the induction principle for identity types. Recall from [section 1.12] that this says that if*

- *for every $x, y : A$ and every $p : x =_A y$ we have a type $D(x, y, p)$, and*
- *for every $a : A$ we have an element $d(a) : D(a, a, \text{refl}_a)$,*

then

- *there exists an element $\text{ind}_{=_A}(D, d, x, y, p) : D(x, y, p)$ for every two elements $x, y : A$ and $p : x =_A y$, such that $\text{ind}_{=_A}(D, d, a, a, \text{refl}_a) \equiv d(a)$.*

The book then reiterates this definition, with basically no natural language, essentially in the raw logical framework devoid of anything but dependent function types.

Definition 7 (Version 2) *In other words, given dependent functions*

$$D : \prod_{(x, y : A)} (x = y) \rightarrow \mathcal{U}$$

$$d : \prod_{a : A} D(a, a, \text{refl}_a)$$

there is a dependent function

$$\text{ind}_{=_A}(D, d) : \prod_{(x, y : A)} \prod_{(p : x = y)} D(x, y, p)$$

such that

$$\text{ind}_{=A}(D, d, a, a, \text{refl}_a) \equiv d(a) \quad (1)$$

for every $a : A$. Usually, every time we apply this induction rule we will either not care about the specific function being defined, or we will immediately give it a different name.

Again, we define this, in Agda, staying as true to the syntax as possible.

```
J : {A : Set}
  → (D : (x y : A) → (x ≡ y) → Set)
  → ((a : A) → (D a a r)) -- → (d : (a : A) → (D a a r))
  → (x y : A)
  → (p : x ≡ y)
  -----
  → D x y p
J D d x .x r = d x
```

It should be noted that, for instance, we can choose to leave out the d label on the third line. Indeed minimizing the amount of dependent typing and using vanilla function types when dependency is not necessary, is generally considered “best practice” Agda, because it will get desugared by the time it typechecks anyways. For the writer of the text; however, it was convenient to define d once, as there are not the same constraints on a mathematician writing in latex. It will again, serve as a nontrivial exercise to deal with when specifying the grammar, and will be dealt with later [ToDo add section]. It is also of note that we choose to include Martin-Löf’s original name J , as this is more common in the computer science literature.

Once the identity type has been defined, it is natural to develop an “equality calculus”, so that we can actually use it in proof’s, as well as develop the higher groupoid structure of types. The first fact, that propositional equality is an equivalence relation, is well motivated by needs of practical theorem proving in Agda and the more homotopically minded mathematician. First, we show the symmetry of equality—that paths are reversible.

Lemma 1 *For every type A and every $x, y : A$ there is a function*

$$(x = y) \rightarrow (y = x)$$

denoted $p \mapsto p^{-1}$, such that $\text{refl}_x^{-1} \equiv \text{refl}_x$ for each $x : A$. We call p^{-1} the **inverse** of p .

Proof 1 (First proof) Assume given $A : \mathcal{U}$, and let $D : \prod_{(x,y:A)} (x = y) \rightarrow \mathcal{U}$ be the type family defined by $D(x, y, p) :\equiv (y = x)$. In other words, D is a function assigning to any $x, y : A$ and $p : x = y$ a type, namely the type $y = x$. Then we have an element

$$d :\equiv \lambda x. \text{refl}_x : \prod_{x:A} D(x, x, \text{refl}_x).$$

Thus, the induction principle for identity types gives us an element $\text{ind}_{=A}(D, d, x, y, p) : (y = x)$ for each $p : (x = y)$. We can now define the desired function $(-)^{-1}$ to be $\lambda p. \text{ind}_{=A}(D, d, x, y, p)$, i.e. we set $p^{-1} :\equiv \text{ind}_{=A}(D, d, x, y, p)$. The conversion rule [missing reference] gives $\text{refl}_x^{-1} \equiv \text{refl}_x$, as required.

The Agda code is certainly more brief:

```

 $\_^{-1} : \{A : \text{Set}\} \{x\ y : A\} \rightarrow x \equiv y \rightarrow y \equiv x$ 
 $\_^{-1} \{A\} \{x\} \{y\} p = \text{J } D\ d\ x\ y\ p$ 
where
   $D : (x\ y : A) \rightarrow x \equiv y \rightarrow \text{Set}$ 
   $D\ x\ y\ p = y \equiv x$ 
   $d : (a : A) \rightarrow D\ a\ a\ r$ 
   $d\ a = r$ 

infixr 50  $\_^{-1}$ 

```

While first encountering induction principles can be scary, they are actually more mechanical than one may think. This is due to the fact that they uniquely compliment the introduction rules of an inductive type, and are simply a means of showing one can “map out”, or derive an arbitrary type dependent on the type which has been inductively defined. The mechanical nature is what allows for Coq’s induction tactic, and perhaps even more elegantly, Agda’s pattern matching capabilities. It is always easier to use pattern matching for the novice Agda programmer, which almost feels like magic. Nonetheless, for completeness sake, the book uses the induction principle for much of Chapter 2. And pattern matching is unique to programming languages, its elegance isn’t matched in the mathematicians’ lexicon.

Here is the same proof via “natural language pattern matching” and Agda pattern matching:

Proof 2 (Second proof) *We want to construct, for each $x, y : A$ and $p : x = y$, an element $p^{-1} : y = x$. By induction, it suffices to do this in the case when y is x and p is refl_x . But in this case, the type $x = y$ of p and the type $y = x$ in which we are trying to construct p^{-1} are both simply $x = x$. Thus, in the “reflexivity case”, we can define refl_x^{-1} to be simply refl_x . The general case then follows by the induction principle, and the conversion rule $\text{refl}_x^{-1} \equiv \text{refl}_x$ is precisely the proof in the reflexivity case that we gave.*

```

 $\_^{-1'} : \{A : \text{Set}\} \{x\ y : A\} \rightarrow x \equiv y \rightarrow y \equiv x$ 
 $\_^{-1'} \{A\} \{x\} \{y\} r = r$ 

```

Next is transitivity-concatenation of paths-and we omit the natural language presentation, which is a slightly more sophisticated argument than for symmetry.

```

 $\_ \bullet \_ : \{A : \text{Set}\} \rightarrow \{x\ y : A\} \rightarrow (p : x \equiv y) \rightarrow \{z : A\} \rightarrow (q : y \equiv z) \rightarrow x \equiv z$ 
 $\_ \bullet \_ \{A\} \{x\} \{y\} p \{z\} q = \text{J } D\ d\ x\ y\ p\ z\ q$ 
where
   $D : (x_1\ y_1 : A) \rightarrow x_1 \equiv y_1 \rightarrow \text{Set}$ 
   $D\ x\ y\ p = (z : A) \rightarrow (q : y \equiv z) \rightarrow x \equiv z$ 
   $d : (z_1 : A) \rightarrow D\ z_1\ z_1\ r$ 
   $d = \lambda\ v\ z\ q \rightarrow q$ 

infixl 40  $\_ \bullet \_$ 

```

Putting on our spectacles, the reflexivity term serves as evidence of a constant path for any given point of any given type. To the category theorist, this makes up the data of an identity map. Likewise, concatenation of paths starts to look like function composition. This,

along with the identity laws and associativity as proven below, gives us the data of a category. And we have not only have a category, but the symmetry allows us to prove all paths are isomorphisms, giving us a groupoid. This isn't a coincidence, it's a very deep and fascinating articulation of power of the machinery we've so far built. The fact the path space over a type naturally must satisfies coherence laws in an even higher path space gives leads to this notion of types as higher groupoids.

As regards the natural language—at this point, the bookkeeping starts to get hairy. Paths between paths, and paths between paths between paths, these ideas start to lose geometric intuition. And the mathematician, while they write

Many of the proofs beyond this point are either routinely made via the induction principle, or even more routinely by just pattern matching on equality paths, we omit the details which can be found in the HoTT book, but it is expected that the GF parser will soon cover such examples.

```

i_l : {A : Set} {x y : A} (p : x ≡ y) → p ≡ r • p
i_l {A} {x} {y} p = J D d x y p
  where
    D : (x y : A) → x ≡ y → Set
    D x y p = p ≡ r • p
    d : (a : A) → D a a r
    d a = r

i_r : {A : Set} {x y : A} (p : x ≡ y) → p ≡ p • r
i_r {A} {x} {y} p = J D d x y p
  where
    D : (x y : A) → x ≡ y → Set
    D x y p = p ≡ p • r
    d : (a : A) → D a a r
    d a = r

leftInverse : {A : Set} {x y : A} (p : x ≡ y) → p-1 • p ≡ r
leftInverse {A} {x} {y} p = J D d x y p
  where
    D : (x y : A) → x ≡ y → Set
    D x y p = p-1 • p ≡ r
    d : (x : A) → D x x r
    d x = r

-- lI : {A : Set} {x y : A} (p : x ≡ y) → p-1 • p ≡ r
-- lI r = r

rightInverse : {A : Set} {x y : A} (p : x ≡ y) → p • p-1 ≡ r
rightInverse {A} {x} {y} p = J D d x y p
  where
    D : (x y : A) → x ≡ y → Set
    D x y p = p • p-1 ≡ r
    d : (a : A) → D a a r
    d a = r

doubleInv : {A : Set} {x y : A} (p : x ≡ y) → p-1 -1 ≡ p
doubleInv {A} {x} {y} p = J D d x y p
  where
    D : (x y : A) → x ≡ y → Set

```

```

D x y p = p-1 -1 ≡ p
d : (a : A) → D a a r
d a = r

```

```

associativity : {A : Set} {x y z w : A} (p : x ≡ y) (q : y ≡ z) (r' : z ≡ w) → p • (q • r') ≡ p • q • r'
associativity {A} {x} {y} {z} {w} p q r' = J D1 d1 x y p z w q r'

```

where

```

D1 : (x y : A) → x ≡ y → Set
D1 x y p = (z w : A) (q : y ≡ z) (r' : z ≡ w) → p • (q • r') ≡ p • q • r'
-- d1 : (x : A) → D1 x x r
-- d1 x z w q r' = r -- why can it infer this
D2 : (x z : A) → x ≡ z → Set
D2 x z q = (w : A) (r' : z ≡ w) → r • (q • r') ≡ r • q • r'
D3 : (x w : A) → x ≡ w → Set
D3 x w r' = r • (r • r') ≡ r • r • r'
d3 : (x : A) → D3 x x r
d3 x = r
d2 : (x : A) → D2 x x r
d2 x w r' = J D3 d3 x w r'
d1 : (x : A) → D1 x x r
d1 x z w q r' = J D2 d2 x z q w r'

```

4 Goals and Challenges

References

- [1] The Univalent foundations program and N.J.) Institute for advanced study (Princeton. *Homotopy Type Theory: Univalent Foundations of Mathematics*. 2013.