

# IDEMPOTENTS IN INTENSIONAL TYPE THEORY

MICHAEL SHULMAN

University of San Diego, 5998 Alcalá Park, San Diego, CA 92110  
*e-mail address:* shulman@sandiego.edu

**ABSTRACT.** We study idempotents in intensional Martin-Löf type theory, and in particular the question of when and whether they split. We show that in the presence of propositional truncation and Voevodsky’s univalence axiom, there exist idempotents that do not split; thus in plain MLTT not all idempotents can be proven to split. On the other hand, assuming only function extensionality, an idempotent can be split if and only if its witness of idempotency satisfies one extra coherence condition. Both proofs are inspired by parallel results of Lurie in higher category theory, showing that ideas from higher category theory and homotopy theory can have applications even in ordinary MLTT.

Finally, we show that although the witness of idempotency can be recovered from a splitting, the one extra coherence condition cannot in general; and we construct “the type of fully coherent idempotents”, by splitting an idempotent on the type of partially coherent ones. Our results have been formally verified in the proof assistant Coq.

## 1. INTRODUCTION

In December 2014 Martín Escardó asked me whether idempotents split in Martin-Löf type theory (MLTT). This paper is a long-winded answer.

Usually, an *idempotent* means a function (necessarily an endofunction) that is equal to its composite with itself,  $f \circ f = f$ . In MLTT, using the propositions-as-types methodology, we might naturally take this to mean a function  $f : X \rightarrow X$ , for some type  $X$ , together with a *witness of idempotency*  $I : \prod_{x:X} (f(f(x)) = f(x))$ , where “=” denotes the identity type. (If we assume function extensionality, as we often will, then to give  $I$  is equivalent to giving  $I' : f \circ f = f$ .)

A *splitting* of an idempotent  $f$  on  $X$  consists of functions  $r : X \rightarrow A$  and  $s : A \rightarrow X$  such that  $r \circ s = \text{id}_A$  and  $s \circ r = f$ . In ZFC set theory, an idempotent always has a splitting with  $A = \{x \in X \mid f(x) = x\}$ , where  $s$  is the inclusion and  $r$  is the corestriction of  $f$ . This

*1998 ACM Subject Classification:* F.4.1, D.1.1.

*Key words and phrases:* Martin-Löf type theory, dependent type theory, idempotent, univalence axiom.

This material is based on research sponsored by The United States Air Force Research Laboratory under agreement number FA9550-15-1-0053. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the United States Air Force Research Laboratory, the U.S. Government, or Carnegie Mellon University.

suggests that in MLTT we ought to consider  $A = \sum_{x:X} (f(x) = x)$ , with  $s$  the first projection  $s(x, p) := x$  and  $r$  defined by  $r(x) := (f(x), I(x))$ . However, as Martín observed, this does not work in general:

**Example 1.1.** Let  $X$  be any type and let  $f = \text{id}_X$  be its identity function, with  $I$  the obvious witness defined by  $I(x) := \text{refl}_x$ . Then with the above-defined  $A$ ,  $s$ , and  $r$ , we have  $\prod_{a:A} (r(s(a)) = a)$  if and only if  $X$  satisfies Uniqueness of Identity Proofs (UIP). (We will prove this in §3.) Since MLTT does not prove that all types satisfy UIP, neither can it prove that this construction always splits an idempotent.

Now, if we were wondering whether MLTT proved some theorem and we had found that the obvious proof of some theorem used a classical axiom such as the law of excluded middle, then it would be natural to seek for counterexamples in nonclassical models (such as topological or realizability models) or disproofs from nonclassical axioms (such as strong Church’s thesis or Brouwerian continuity principles). Similarly, having found that the obvious way to split idempotents depends on UIP, it is natural to seek counterexamples in models that violate UIP or disproofs from axioms that contradict it.

This leads us into the recently discovered realm of Homotopy Type Theory and Univalent Foundations [APW13, Awo12, PW14, Uni13]. Models which violate UIP, such as the Hofmann–Streicher groupoid model [HS98] and Voevodsky’s simplicial set model [KLV12], tend to be based on the idea that types are *homotopy spaces* or  $\infty$ -*groupoids*. The principal known axiom that contradicts UIP — Voevodsky’s univalence axiom — is also based on this idea.

This suggests that when seeking inspiration from classical mathematics, instead of ZFC set theory we should look to homotopy theory and  $\infty$ -groupoid theory. In these fields, an important role is played by *homotopy coherence*. When a structure satisfies some property “up to homotopy”, for many purposes it is not enough to simply have such a homotopy; often one requires this homotopy to satisfy some natural axiom(s) at the next dimension up — and that only up to homotopy, a homotopy that in turn satisfies its own axioms up to even higher homotopy, and so on to infinity.

For instance, instead of a *group* we may consider an  $\infty$ -*group* (a.k.a. “grouplike  $A_\infty$ -space”). This is a space  $X$  with a multiplication  $m : X \times X \rightarrow X$  that is, among other things, associative but only up to homotopy: for any  $x, y, z \in X$  instead of  $m(m(x, y), z)$  and  $m(x, m(y, z))$  being equal, they are connected by a path depending continuously on  $x, y, z$ . These paths are then required to satisfy a further property: for any  $x, y, z, w \in X$  there is a pentagon that can be built out of these paths, and we require that there be a continuous way to “fill in” that pentagon inside  $X$ . From those filled pentagons one can then construct the boundary of a certain polyhedron, which we require to have a continuous filler, and so on. If we stop at any finite stage, we obtain a much more poorly-behaved notion.

Now, under the homotopical interpretation of MLTT, a witness of idempotency  $I : \prod_{x:X} (f(f(x)) = f(x))$  corresponds to a *homotopy* from  $f \circ f$  to  $f$ . Thus, from a homotopy-theoretic point of view, it is natural to expect that  $I$  itself would not be enough to obtain a well-behaved notion of “idempotent” (such as, for instance, one that can be split): we should ask it to satisfy a further property analogous to filling the pentagon, and that filler should itself satisfy a higher axiom up to homotopy, and so on. In the context of  $\infty$ -categories, such a definition of *fully-coherent idempotent* has been given by Lurie in [Lur09, §4.4.5], along with proofs that every fully-coherent idempotent splits and every split idempotent is fully-coherent.

Unfortunately, there is a well-known problem with representing such fully-coherent structures in MLTT: on the face of it they seem to require a tower of infinitely many terms, each dependent on the previous ones, which is not something that can be defined as a single object in MLTT. This is somewhat disheartening for the project of splitting idempotents. However, it's important not to read more into the results of Lurie cited above than they say. They do say that if a function  $f$  can be written as  $s \circ r$  where  $r \circ s = \text{id}$ , then  $f$  admits a “coherent system of idempotence data”. They *don't* say that if  $f$  is idempotent with a specified homotopy  $I$ , and  $f$  splits, then  $I$  must *itself* admit an extension to a coherent system of idempotence data. Therefore:

- Even though a split idempotent automatically gives rise to an infinite system of coherence data, it doesn't follow that in order to *construct* a splitting we would necessarily need to *give* an infinite system of coherence.
- It's not too hard to give examples of homotopies  $I$  that are not coherent, but it's rather less obvious how to give an example of an incoherent idempotent for which there doesn't exist some *other* homotopy that is coherent.

Fortunately, Lurie has already addressed these questions as well (still in the  $\infty$ -categorical context). In [Lur14, Warning 1.2.4.8] he gave an example of an incoherent idempotent that does *not* split, and in [Lur14, Lemma 7.3.5.14] he showed that to construct a splitting, *one* additional coherence datum suffices. Inspired by these results, we will set out to transfer them to MLTT, as follows:

- (1) Assuming propositional truncation and the univalence axiom, we can adapt Lurie's counterexample to show that a single witness of idempotency  $I : \prod_{x:X} (f(f(x)) = f(x))$  is insufficient to construct a splitting. In fact, our construction is slightly simpler than Lurie's, and involves an object familiar to constructive mathematicians: the Cantor space  $2^{\mathbb{N}}$ .
- (2) However, under the weaker assumption of function extensionality, we can adapt Lurie's construction to show that if we also have  $J : \prod_{x:X} (\text{ap}_f(I(x)) = I(f(x)))$ , then we can construct a splitting. (Here  $\text{ap}_f$  denotes the action of  $f$  on witnesses of equality; sometimes it is called *resp.*) Our construction is actually the dual of Lurie's: we use a limit where he uses a colimit. A colimit would probably also work, but would require further assumptions on type theory for its construction and well-behavedness.

Note that the latter positive result does *not* require the univalence axiom. So although inspired by higher category theory, we obtain a result that should be of interest even in pure intensional MLTT.

Based on these results, we propose that, as in higher category theory, the unadorned word *idempotent* should not be used for the “incoherent” notion that includes only a single witness  $I$ . Instead we will call the pair  $(f, I)$  a **pre-idempotent**.

One might think that the triple  $(f, I, J)$  ought to deserve the name “idempotent”, since although it does not include all the higher coherence data, we have seen that it does suffice to construct a splitting. However, this is not the case. It is true, in the  $\infty$ -categorical world, that a splitting induces a fully coherent idempotent in Lurie's sense, and hence so does a triple  $(f, I, J)$ ; but nothing guarantees that the resulting coherent idempotent is an extension of  $(f, I, J)$  itself. In fact, we will show in MLTT that it is an extension of  $(f, I)$ , but not in general of  $J$ : assuming univalence and propositional truncation, there exist choices for  $J$  that are not coherentifiable at all. For these reasons, we will instead call a

triple  $(f, I, J)$  a **quasi-idempotent**; it is analogous to the “incoherent, but coherentifiable, equivalences” that in [Uni13] are called *quasi-inverses*.

At this point we may wonder whether there is *any* way to define the word “idempotent” in MLTT in a way that will translate to the correct notion homotopically. There is one answer that is somewhat “cheap”: by [Lur09, Corollary 4.4.5.14], in an  $\infty$ -category the space of idempotents on an object  $X$  is equivalent to the space of *retractions* of  $X$ , meaning quadruples  $(A, r, s, H)$  where  $r : X \rightarrow A$  and  $s : A \rightarrow X$  and  $H$  is a homotopy  $r \circ s \sim \text{id}_A$ . The latter can be defined in MLTT (with universes) as

$$\text{Retr}(X) := \sum_{A:\text{Type}} \sum_{r:X \rightarrow A} \sum_{s:A \rightarrow X} \prod_{a:A} (r(s(a)) = a),$$

and if we assume the univalence axiom, it will have the correct homotopy type. Thus, we could define an *idempotent* on  $X$  to be an inhabitant of this type.

Of course, this would be rather unsatisfying: we expect an “idempotent” to consist of a map  $f : X \rightarrow X$  equipped with some kind of structure, and we expect the construction of its splitting to be a nonvacuous operation. Moreover, it has an actual technical drawback as well: since it involves a sum over a universe  $\text{Type}$ , it lives in a universe one higher than that of  $X$ .

Both of these problems can be solved with the following trick. If we define the type of quasi-idempotents in the expected way:

$$\text{QIdem}(X) := \sum_{f:X \rightarrow X} \sum_{I:\prod_{x:X} (f(f(x))=f(x))} \prod_{x:X} (\text{ap}_f(I(x)) = I(f(x)))$$

then the above splitting construction yields a map

$$\text{split} : \text{QIdem}(X) \rightarrow \text{Retr}(X).$$

On the other hand, since every retraction induces a coherent idempotent, we have a map

$$\text{uli} : \text{Retr}(X) \rightarrow \text{QIdem}(X)$$

and these two maps can be shown to exhibit  $\text{Retr}(X)$  itself as a retract of  $\text{QIdem}(X)$ . Therefore, the composite  $\text{uli} \circ \text{split}$  is a quasi-idempotent on  $\text{QIdem}(X)$ , and if we construct its splitting as above, we obtain a type equivalent to  $\text{Retr}(X)$ . This splitting type is what we propose as the definition of **(fully coherent) idempotent**: it has the correct homotopy type; it is by construction an equipping of an endomap with data (indeed, infinitely many data, encoded internally by way of the natural numbers type); and it lies in the same universe as  $X$ .

The plan of the paper is as follows. In §2 we recall some notation and terminology from [Uni13]. In §3 we ease into the study of idempotents by considering several hypotheses (due to Martín Escardó) under which pre-idempotents can be split. The next two sections contain the main results: in §4 we give our example of a pre-idempotent that admits no splitting (assuming propositional truncation and the univalence axiom), and in §5 we construct a splitting of any quasi-idempotent (assuming function extensionality).

The remaining sections are concerned with the more technical coherence questions. In §6 we show that  $\text{split}$  exhibits  $\text{Retr}(X)$  as a retract of  $\text{QIdem}(X)$ . In §7 we show that this retraction is not an equivalence, and conclude that although the underlying pre-idempotent of a quasi-idempotent can be recovered from its splitting, the coherence datum  $J$  cannot in general be. In §8 we complete a proof from §7 that requires a lengthy analysis of some classifying spaces in type theory. Finally, in §9 we define the type of fully-coherent idempotents, and in §10 we conclude with some remaining open problems.

Throughout, we will argue in the informal style of [Uni13], and we will make use of the basic results from Chapters 1–4 thereof. However, all the main results of this paper have also been formally verified in the proof assistant Coq, using the Homotopy Type Theory library [HoT15], and are available as part of that library. As of the date of publication, the correspondence between sections of this paper and files in the library is:

Section	Library File
§3	Idempotents.v
§4	Spaces/BAut/Cantor.v
§§5–7	Idempotents.v
§8	Spaces/BAut.v and Spaces/BAut/Bool.v
§9	Idempotents.v

The `idempotents-paper` git tag records this version of the library.

## 2. SOME NOTATION AND TERMINOLOGY

For the most part, we adopt the notation and terminology of [Uni13]. We write  $\prod_{x:A} B(x)$  and  $\sum_{x:A} B(x)$  for dependent product and sum as usual in MLTT, with their non-dependent special cases  $A \rightarrow B$  and  $A \times B$ . We write the identity type of two elements  $x, y : A$  as  $x =_A y$ , or usually just  $x = y$ ; its canonical elements are  $\text{refl}_x : x = x$ . We write  $x \equiv y$  for a judgmental equality, and  $a \equiv b$  if  $a$  is currently being defined to equal  $b$ .

A type  $A$  is called a **mere proposition** if we have  $\prod_{x,y:A} (x = y)$ . In other words,  $\text{isprop}(A) := \prod_{x,y:A} (x = y)$ . It is said to **be a set**, or to **satisfy** UIP (Uniqueness of Identity Proofs), if  $\prod_{x,y:A} \text{isprop}(x = y)$ , or equivalently  $\prod_{x:X} \prod_{p:x=x} (p = \text{refl}_x)$ .

For functions  $f, g : A \rightarrow B$ , we write  $f \sim g$  for the type  $\prod_{x:A} (f(x) = g(x))$ , and call it the type of **homotopies** from  $f$  to  $g$ . The **function extensionality** axiom, which we will almost always have available (either by explicit assumption, or as a consequence of some other assumption), says that this type is equivalent (see below) to the identity type  $f =_{A \rightarrow B} g$ .

For types  $A$  and  $B$ , we write  $A \simeq B$  for the type of **equivalences** from  $A$  to  $B$ . This is defined as  $\sum_{f:A \rightarrow B} \text{isequiv}(f)$ , where  $\text{isequiv}(f)$  is any one of a number of well-behaved definitions, the first of which was due to Voevodsky; see [Uni13, Chapter 4] for details. The important properties are that  $\text{isequiv}(f)$  if and only if  $\sum_{g:B \rightarrow A} (f \circ g \sim \text{id}) \times (g \circ f \sim \text{id})$  (we generally use the “if” direction of this to construct equivalences), and that  $\text{isequiv}(f)$  is a mere proposition. There is a canonical map  $(A = B) \rightarrow (A \simeq B)$ , and Voevodsky’s **univalence axiom** says that this map is itself an equivalence.

The **propositional truncation** is, when assumed, a rule associating to every type  $A$  a type  $\|A\|$  which is a mere proposition, and a map  $|-| : A \rightarrow \|A\|$ , such that any function from  $A$  to a mere proposition factors judgmentally through  $\|A\|$ . In other words, if  $B$  is a mere proposition and  $f : A \rightarrow B$ , then there exists  $g : \|A\| \rightarrow B$  such that  $f(a) \equiv g(|a|)$  for all  $a : A$ . We sometimes pronounce  $\|A\|$  as “**merely**  $A$ ”, e.g. if we have an element of  $\|A \simeq B\|$  we say that  $A$  and  $B$  are merely equivalent. Both univalence and propositional truncation imply function extensionality; the former is due to Voevodsky (see e.g. [Uni13, §4.9]) and the latter to [KECA14, Corollary 9.3].

With homotopy-theoretic intuition in mind, elements of identity types (i.e. witnesses of equality) are sometimes called **paths**. For  $p : x =_A y$  and  $q : y =_A z$ , we have  $p \cdot q : x =_A z$  (a witness of transitivity) and  $p^{-1} : y =_A x$  (a witness of symmetry), defined using the eliminator of the identity type. Similarly, if  $f : A \rightarrow B$ , we have  $\text{ap}_f(p) : f(x) =_B f(y)$ , and

this operation is functorial (up to propositional equality) in two ways:  $\mathbf{ap}_g \circ \mathbf{ap}_f = \mathbf{ap}_{g \circ f}$ , and  $\mathbf{ap}_f(p \cdot q) = \mathbf{ap}_f(p) \cdot \mathbf{ap}_f(q)$ .

We will frequently use the fact that homotopies between functions satisfy a *naturality* property [Uni13, Lemma 2.4.3]. Specifically, given  $g, h : B \rightarrow C$  and  $L : g \sim h$ , for any  $b_1, b_2 : B$  and  $p : b_1 = b_2$ , we have

$$\mathbf{ap}_g(p) \cdot L(b_2) = L(b_1) \cdot \mathbf{ap}_h(p).$$

The other important facts we will use from [Uni13] are the theorems from its Chapter 2 that characterize the identity types of different type formers (sometimes requiring univalence and function extensionality). For instance, the type  $(a_1, b_1) =_{A \times B} (a_2, b_2)$  is equivalent to  $(a_1 =_A a_2) \times (b_1 =_B b_2)$ , i.e. two ordered pairs are equal just when their components are. In most cases these results are reasonably intuitive.

### 3. SOME PRE-IDEMPOTENTS THAT SPLIT

As suggested in the introduction, we define:

**Definition 3.1.** A **pre-idempotent** is an endofunction  $f : X \rightarrow X$  equipped with a witness of idempotency  $I : f \circ f \sim f$ .

**Definition 3.2.** A **retract** of a type  $X$  consists of a type  $A$ , functions  $s : A \rightarrow X$  and  $r : X \rightarrow A$ , and a homotopy  $H : r \circ s \sim \text{id}_A$ . A **splitting** of an endofunction  $f : X \rightarrow X$  is a retraction  $(A, r, s, H)$  together with a homotopy  $K : s \circ r \sim f$ .

The following is fairly obvious.

**Lemma 3.3.** *If  $f$  has a splitting, then it is pre-idempotent.*

*Proof.* Clearly anything homotopic to a pre-idempotent is pre-idempotent, so it suffices to show that if we have a retraction  $(A, r, s, H)$  then  $s \circ r$  is pre-idempotent. In this case, for any  $x : X$ , we can define  $I(x) := \mathbf{ap}_s(H(r(x))) : s(r(s(r(x)))) = s(r(x))$ .  $\square$

We can also show easily that splittings are essentially unique in at least a weak sense.

**Lemma 3.4.** *Suppose  $f : X \rightarrow X$  has two splittings  $(A, s, r, H, K)$  and  $(A', s', r', H', K')$ . Then  $A \simeq A'$ .*

*Proof.* We have two functions  $r's : A \rightarrow A'$  and  $rs' : A' \rightarrow A$ , and their composites are homotopic to identities:

$$r'srs' \sim r'fs' \sim r's'r's' \sim \text{id}_{A'}$$

and similarly  $rs'r's \sim \text{id}_A$ .  $\square$

We expect that a split endofunction is not only pre-idempotent, but fully-coherently idempotent. As remarked in the introduction, it is difficult to define fully-coherent idempotents in type theory, but we can at least define the next step of coherence.

**Definition 3.5.** A **quasi-idempotent** is a pre-idempotent  $(f, I)$  together with a witness of coherence  $J : \prod_{x:X} (\mathbf{ap}_f(I(x)) = I(f(x)))$ .

**Lemma 3.6.** *If  $f$  has a splitting, then it is quasi-idempotent.*

*Proof.* As in Lemma 3.3, it suffices to show that for any retraction  $(A, s, r, H)$ ,  $s \circ r$  is quasi-idempotent. For this case, in Lemma 3.3 we defined  $I(x) := \mathbf{ap}_s(H(r(x))) : s(r(s(r(x)))) =$

$s(r(x))$ . Thus, for  $x : X$  the desired type of  $J(x)$  is

$$\mathbf{ap}_f(\mathbf{ap}_s(H(r(x)))) = \mathbf{ap}_s(H(r(f(x)))).$$

This is equivalent to

$$\mathbf{ap}_s(\mathbf{ap}_{r \circ s}(H(r(x)))) = \mathbf{ap}_s(H(r(s(r(x))))).$$

Peeling off an  $\mathbf{ap}_s$ , and letting  $a := r(x)$ , it will suffice to show that for any  $a : A$  we have

$$\mathbf{ap}_{r \circ s}(H(a)) = H(r(s(a))).$$

At first this seems like a nontrivial property of  $H$ . However, in fact it is automatic. For by naturality of the homotopy  $H$  applied at the equality  $H(a)$ , we have

$$\mathbf{ap}_{r \circ s}(H(a)) \cdot H(a) = H(r(s(a))) \cdot H(a).$$

Now we can cancel  $H(a)$  from both sides to obtain the desired result.  $\square$

We now give a few conditions under which pre-idempotents can be split. Our first observation is:

**Theorem 3.7.** *If  $X$  is a set, then any pre-idempotent on  $X$  has a splitting.*

*Proof.* Define  $A := \sum_{x:X} (f(x) = x)$ , and let  $s$  and  $r$  be defined by  $s(x, p) = x$  and  $r(x) = (f(x), I(x))$ . Now for  $x : X$ , we have  $s(r(x)) \equiv f(x)$  by definition; hence we can take  $K(x) := \mathbf{refl}_{f(x)}$ . On the other hand, for  $(x, p) : A$  we have  $r(s(x, p)) \equiv (f(x), I(x))$ ; thus  $H(x, p)$  must inhabit  $((f(x), I(x)) = (x, p))$ . By [Uni13, Theorems 2.7.2 and 2.11.3], to give an element of this type we must give  $q : f(x) = x$  and  $r : \mathbf{ap}_f(q)^{-1} \cdot I(x) \cdot q = p$ . But we can take  $q := p$ , and obtain  $r$  from the assumption that  $X$  is a set.  $\square$

Now here is our elaboration of Example 1.1, showing that this construction cannot always work.

**Example 3.8** (Escardó). Let  $X$  be any type and  $f := \mathbf{id}_X$ , with  $I(x) := \mathbf{refl}_x$ . Then with the above-defined  $A$ ,  $s$ , and  $r$ , the desired type of  $H(x, p)$  is equivalent to  $q^{-1} \cdot \mathbf{refl}_x \cdot q = p$ , and hence to  $\mathbf{refl}_x = p$ . If this is true for all  $x : X$  and all  $p : x = x$ , then  $X$  satisfies UIP.

Escardó has also observed a couple of other situations in which pre-idempotents can be split. For the first, recall from [KECA14] that a function  $f : X \rightarrow Y$  is **weakly constant** if we have a witness  $\prod_{x,y:X} (f(x) = f(y))$ .

**Theorem 3.9** (Escardó). *If a pre-idempotent is weakly constant, then it has a splitting.*

*Proof.* We use the same construction as in Theorem 3.7; by following the proof thereof, it remains only to construct  $H$ . However, by [KECA14, Lemma 4.1]<sup>1</sup>, when  $f$  is weakly constant, our type  $A := \sum_{x:X} (f(x) = x)$  (there called  $\mathbf{fix}(f)$ ) is a mere proposition, i.e. we have  $\prod_{a,b:A} (a = b)$ . This makes the construction of  $H$  trivial.  $\square$

Conversely, it is easy to see that if an endofunction splits through a mere proposition, then it is weakly constant.

For the second, recall from [KECA14] that a type is called **collapsible** if it admits a weakly constant endofunction. This is equivalent to the existence of some mere proposition  $P$  with functions  $A \rightarrow P$  and  $P \rightarrow A$  (since any function that factors through a mere proposition is weakly constant, while by [KECA14, Lemma 4.1] if  $f$  is weakly constant we can take  $P := \mathbf{fix}(f)$ ). Moreover, if we have propositional truncation, collapsibility is

<sup>1</sup>Also formalized in [HoT15] as `ishprop_fix_wconst` in `Constant.v`.

equivalent to the existence of a map  $\|A\| \rightarrow A$ , a property which one may call having **split support**. Finally, recall from [Uni13, Lemma 7.6.2] that a function  $f : A \rightarrow B$  is said to be an **embedding** if for all  $b : B$  the type  $\sum_{a:A}(f(a) = b)$  is a mere proposition.

The following theorem is our first example of a definable splitting in which the splitting type  $A$  is *not* the obvious  $\sum_{x:X}(f(x) = x)$ .

**Theorem 3.10** (Escardó). *An endofunction  $f$  has a splitting in which the section  $s$  is an embedding if and only if it is pre-idempotent and the type  $f(x) = x$  is collapsible for all  $x$ .*

(It is arguably more natural to formulate this theorem in terms of split support. The advantage of using collapsibility instead is that it makes sense even in the absence of propositional truncation.)

*Proof.* First suppose  $f$  is pre-idempotent and each  $f(x) = x$  is collapsible. Thus, for each  $x$  there is a mere proposition  $P_x$  and maps  $u_x : (f(x) = x) \rightarrow P_x$  and  $v_x : P_x \rightarrow (f(x) = x)$ . (If we have propositional truncation, we can take  $P_x := \|f(x) = x\|$ , and the reader may find it easier to think about this case.) We define  $A := \sum_{x:X} P_x$ , with  $s(x, p) := x$  and  $r(x) := (f(x), u_{f(x)}(I(x)))$ , while  $K(x) := \text{refl}_{f(x)}$  as before. For  $H$ , given  $(x, p) : A$  where  $x : X$  and  $p : P_x$ , we must show that  $(f(x), u_{f(x)}(I(x))) = (x, p)$ , which as before amounts to giving  $q : f(x) = x$  and an equality  $q_*(u_{f(x)}(I(x))) = p$ , where  $q_* : P_{f(x)} \rightarrow P_x$  denotes transport along  $q$ . But we can define  $q := v_x(p)$ , while the remaining equality is trivial since  $P_x$  is a mere proposition.

Now conversely, suppose  $f$  has a splitting in which  $s$  is an embedding; it remains to show that  $f(x) = x$  is collapsible for all  $x : X$ . Since  $s$  is an embedding, the type  $\sum_{a:A}(s(a) = x)$  is a mere proposition. Thus, it will suffice to construct maps in both directions relating this type to  $f(x) = x$ , or equivalently to the type  $s(r(x)) = x$ . In one direction, given  $p : s(r(x)) = x$ , we have  $(r(x), p) : \sum_{a:A}(s(a) = x)$ . In the other, given  $a : A$  and  $p : s(a) = x$ , we have  $s(r(x)) = s(r(s(a))) = s(a) = x$ .  $\square$

**Remarks 3.11.**

- (1) Theorem 3.7 is a special case of Theorem 3.10: if  $X$  is a set, then each type  $f(x) = x$  is a mere proposition, hence trivially collapsible. Moreover, by [KECA14, Theorem 3.10], if *every* type  $x =_X y$  is collapsible, then  $X$  is necessarily a set. However, there do exist functions on non-sets to which Theorem 3.10 applies; a trivial example is  $X := Y + \mathbf{1}$  with  $f(x) := \text{inr}(\text{tt})$ .
- (2) On the other hand, there can exist retractions for which the section is not an embedding. For instance, any map  $\mathbf{1} \rightarrow X$  exhibits  $\mathbf{1}$  as a retract of  $X$ , but to say that all such maps are embeddings is just to say that  $X$  is a set. Thus, Theorem 3.10 emphasizes another way in which idempotents in homotopy theory differ from idempotents in set theory, since in set theory the splitting of an idempotent *always* injects into the original set.
- (3) When the conditions of Theorem 3.10 fail, it doesn't generally mean there is any particular  $x : X$  such that  $f(x) = x$  is not collapsible. It only means we cannot assert that " $f(x) = x$  is collapsible for all  $x : X$ ", since such an assertion would imply an impossible "naturalness" of the collapsing maps.
- (4) Perhaps surprisingly, none of the results in this section require even function extensionality.



## 4. A PRE-IDEMPOTENT THAT DOESN'T SPLIT

As mentioned in the introduction, it's easy to give examples of idempotence witnesses  $I : f \circ f \sim f$  that cannot be extended to a coherent system of idempotence data.

**Example 4.1.** Let  $X$  be any type with a point  $x_0 : X$  for which there exists a nontrivial  $p : x_0 = x_0$ , i.e. such that  $p \neq \text{refl}_{x_0}$ . (For instance, in the presence of the univalence axiom, we could let  $X$  be the universe, with  $x_0$  a type admitting a nonidentity self-equivalence, such as **2**.) Define  $f : X \rightarrow X$  by  $f(x) := x_0$  for all  $x$ , and let  $I(x) := p$  for all  $x$ . Then  $(f, I)$  is a pre-idempotent. But  $\text{ap}_f(q) = \text{refl}_{x_0}$  for all  $q$ , so the second-level coherence type  $\prod_{x:X}(\text{ap}_f(I(x)) = I(f(x)))$  is equivalent to  $\prod_{x:X}(\text{refl}_{x_0} = p)$ , which by assumption is not inhabited.

However, it's less clear how to exhibit a pre-idempotent  $(f, I)$  for which there cannot exist any *other* witness  $I'$  that *is* coherent. For instance, in the above example, we could simply have taken  $I'(x) := \text{refl}_{x_0}$ .

We will describe an example inspired by that of [Lur14, Warning 1.2.4.8], but not quite identical to it. In Lurie's example, the space  $X$  is the classifying space of the group of endpoint-preserving self-homeomorphisms of the unit interval  $[0, 1]$ . However, the essential feature of this choice, for the purposes of the example, is that two such homeomorphisms can be shrunk by a factor of 2 and glued together to form a new such. This is reminiscent of Freyd's universal characterization of  $[0, 1]$  (see e.g. [Joh02, D4.7.17]), but in fact it can be completely divorced from the topology. Thus, we will instead use a type familiar to constructive mathematicians: the Cantor space.

**Definition 4.2.** The **Cantor space** is the type  $C := (\mathbb{N} \rightarrow \mathbf{2})$ .

The essential property of  $C$ , for our purposes, is the following.

**Lemma 4.3.** *Assuming function extensionality,  $C \simeq (C + C)$ .*

*Proof.* From left to right, given  $c : \mathbb{N} \rightarrow \mathbf{2}$  we define  $c'(n) := c(n + 1)$ , and split into cases based on whether  $c(0)$  is 0 or 1. In the former case, we send  $c$  to  $\text{inl}(c')$ , and in the second case we send it to  $\text{inr}(c')$ .

From right to left, we send  $\text{inl}(c)$  to  $c_0$ , where  $c_0(0) := 0$  and  $c_0(n + 1) := c(n)$ ; and similarly we send  $\text{inr}(c)$  to  $c_1$  where  $c_1(0) := 1$  and  $c_1(n + 1) := c(n)$ . It is easy to check that these are inverse equivalences.  $\square$

We now consider the “classifying space of the automorphism group of  $C$ ”, starting by defining it.

**Assumption 4.4.** *For the rest of this section we assume both univalence and propositional truncation.*

From this assumption we also get function extensionality. In fact, as noted earlier, both univalence and propositional truncation separately imply it.

**Definition 4.5.** For any  $Y : \text{Type}$ , we define  $\text{BAut}(Y) := \sum_{Z:\text{Type}} \|Z = Y\|$ .

Because  $\|Z = Y\|$  is a mere proposition, if we have  $(Z, e)$  and  $(Z', e')$  in  $\text{BAut}(Y)$ , the type  $(Z, e) = (Z', e')$  is equivalent to  $Z = Z'$  and hence (by univalence) to  $Z \simeq Z'$ . This justifies abusing the notation by identifying an element of  $\text{BAut}(Y)$  with its first component, which is a type  $Z$  that comes equipped with an element of  $\|Z = Y\|$ .

In particular, we have the canonical element  $(Y, |\text{refl}_Y|) : \mathbf{BAut}(Y)$ , and the type  $(Y, |\text{refl}_Y|) = (Y, |\text{refl}_Y|)$  (the “loop space” of  $\mathbf{BAut}(Y)$  at this “basepoint”) is equivalent to  $Y \simeq Y$ , the type of automorphisms of  $Y$ . It is in this sense that  $\mathbf{BAut}(Y)$  is a classifying space for the automorphism group of  $Y$ . (The type  $\mathbf{BAut}(Y)$  is also a classifying space in another sense: to give a map  $A \rightarrow \mathbf{BAut}(Y)$  is equivalent to giving a map  $p : B \rightarrow A$  such that every fiber is merely equivalent to  $Y$ , i.e. for each  $a : A$  we have  $\|Y = \sum_{b:B} (p(b) = a)\|$ .)

Our incoherent pre-idempotent will live on the type  $X := \mathbf{BAut}(C)$ , where  $C$  is the Cantor space. Thus, we must next construct a particular map  $f : X \rightarrow X$ . If we translated Lurie’s construction directly, we would do this by first defining an automorphism  $F$  of the group  $\mathbf{Aut}(C)$ , by sending an automorphism  $h$  to the automorphism  $F(h)$  defined as the composite

$$C \simeq C + C \xrightarrow{h+\text{id}} C + C \simeq C$$

where the equivalences come from Lemma 4.3. Then we would use the fact that an automorphism of a group induces an automorphism of its classifying space to obtain from  $F$  an automorphism of  $\mathbf{BAut}(Y)$ .

However, although this fact is standard in homotopy theory, it is not obvious from our definition of  $\mathbf{BAut}(Y)$  in type theory that any automorphism of the group  $\mathbf{Aut}(Y)$  induces an automorphism of the type  $\mathbf{BAut}(Y)$ . It can be deduced from the alternative construction of classifying spaces in [LF14]; but fortunately in our case there is a better approach.

The univalence axiom has allowed us to define  $\mathbf{BAut}(Y)$  in such a way that its elements literally *are* types that are merely equivalent to  $Y$  (or more precisely, equipped with such a mere equivalence). Thus, we can define  $f$  to act directly on such types, rather than indirectly on their automorphisms. Specifically, if we define  $f(Z) := Z + C$ , then the *induced* action on automorphisms will automatically have the intended effect as shown above.

All we have to do is verify that this definition indeed defines an endomorphism of  $\mathbf{BAut}(C)$ , i.e. that if  $\|Z = C\|$  then also  $\|Z + C = C\|$ . By the induction principle of propositional truncation, it suffices to prove that if  $Z = C$  then  $Z + C = C$ , and by univalence it suffices to prove that if  $e : Z \simeq C$  then  $Z + C \simeq C$ . But for this we have the composite

$$Z + C \simeq C + C \simeq C$$

where the first equivalence is  $e + \text{id}$  and the second is Lemma 4.3.

Next, we have to construct a witness of pre-idempotency for  $f$ , i.e. we must show that for any  $Z : \mathbf{BAut}(C)$  we have  $(Z + C) + C = Z + C$ . We again apply univalence and then use the following composite equivalence:

$$(Z + C) + C \simeq Z + (C + C) \simeq Z + C \tag{4.1}$$

consisting of the associativity of coproducts together with Lemma 4.3.

We are now ready for the central theorem of this section.

**Theorem 4.6.** *There exists a pre-idempotent on  $X := \mathbf{BAut}(C)$  that does not split.*

*Proof.* The construction of the pre-idempotent is as above; it remains to show that  $f$  does not split. Lurie’s argument is that if it did, then the colimit  $X \xrightarrow{f} X \xrightarrow{f} X \xrightarrow{f} \dots$  would be its splitting, and hence the map from  $X$  to that colimit would be surjective on fundamental groups; whereas  $f$  itself is certainly not surjective on fundamental groups and so this is impossible. In type theory, colimits are difficult to work with, though homotopy type theory with higher inductive types makes them more tractable than otherwise. However,

we can fortunately again give a more direct argument, based on our concrete construction of  $B\text{Aut}(C)$ .

Suppose for contradiction that  $f$  is split. Then by Lemma 3.6 it is quasi-idempotent, with witnesses  $I$  and  $J$ . For any  $Z : B\text{Aut}(C)$ , we have

$$J(Z) : \mathbf{ap}_f(I(Z)) =_{f(f(f(Z)))=f(f(Z))} I(f(Z)).$$

Since  $f(f(f(Z))) = Z + C + C + C$  and  $f(f(Z)) = Z + C + C$ , by univalence and function extensionality,  $J(Z)$  may equivalently be regarded as a homotopy between two specified equivalences  $(Z + C + C + C) \rightarrow (Z + C + C)$ .

The first of these equivalences (corresponding to  $\mathbf{ap}_f(I(Z))$ ) decomposes the domain and codomain as  $(Z + C + C) + C$  and  $(Z + C) + C$ , mapping the first summand  $Z + C + C$  to  $Z + C$  by  $I(Z)$  and the second summand  $C$  to  $C$  by the identity. As for the second equivalence, if  $I$  were the witness (4.1) that we gave above, then the equivalence  $Z + C + C + C \rightarrow Z + C + C$  corresponding to  $I(f(Z))$  would bracket the domain instead as  $(Z + C) + (C + C)$  and map it to  $(Z + C) + C$  by the identity on  $Z + C$  and the “fold” equivalence  $C + C \rightarrow C$ . Thus, the two could not possibly be homotopic, since they would send the third summand of  $Z + C + C + C$  to different summands of the codomain.

This argument doesn’t quite work as stated, since  $I$  might *not* be the same proof of idempotency that we gave above. (Remember that we are supposing only that  $f$  is split, hence quasi-idempotent, in *some* way, since the claim to prove is that  $f$  is not split, which makes no reference to any previously existing witness of pre-idempotence.) However, whatever  $I$  is, it is defined “for all  $Z : B\text{Aut}(C)$ ”. This implies that the induced equivalences  $Z + C + C \rightarrow Z + C$  must be *natural* with respect to equivalences between  $Z$ s (this is not exactly the same sort of naturality that we mentioned in §2, but it follows similarly). In other words, for any  $Z, Z' : B\text{Aut}(C)$  and equivalence  $e : Z \simeq Z'$ , the following square must commute (up to homotopy):

$$\begin{array}{ccc} Z + C + C & \xrightarrow{I(Z)} & Z + C \\ e + \text{id} + \text{id} \downarrow & & \downarrow e + \text{id} \\ Z' + C + C & \xrightarrow{I(Z')} & Z' + C \end{array} \quad (4.2)$$

In particular, we can take  $Z$  and  $Z'$  to be both  $f(C)$ , i.e.  $C + C$ , and let  $e$  be the “flip” automorphism  $C + C \simeq C + C$  that interchanges the summands. Then the horizontal maps in (4.2) are both the equivalence  $(C + C) + C + C \rightarrow (C + C) + C$  induced by  $I(f(C))$ , and (4.2) itself becomes

$$\begin{array}{ccc} C + C + C + C & \xrightarrow{I(f(C))} & C + C + C \\ e + \text{id} + \text{id} \downarrow & & \downarrow e + \text{id} \\ C + C + C + C & \xrightarrow{I(f(C))} & C + C + C \end{array} \quad (4.3)$$

Consider elements of the third and fourth summands in the upper-left corner, which are fixed by  $e + \text{id} + \text{id}$  on the left. Since the two horizontal maps are both  $I(f(C))$ , it must be that the image of any such element under  $I(f(C))$  is fixed by  $e + \text{id}$  on the right. But the only elements of  $C + C + C$  fixed by  $e + \text{id}$  are those in the third summand. Thus,  $I(f(C))$  must map the last two summands in the domain to the last one summand in the

codomain, just as our original witness of pre-idempotency did, so our previous argument to a contradiction kicks in.  $\square$

Note that we actually showed a bit more: there is a pre-idempotent on  $B\text{Aut}(C)$  that is not even quasi-idempotent. Because univalence and propositional truncation are consistent assumptions, we conclude:

**Corollary 4.7.** *It is impossible to prove in MLTT that all pre-idempotents split, or even that all pre-idempotents are quasi-idempotent.*  $\square$

## 5. ALL QUASI-IDEMPOTENTS SPLIT

We now show, in contrast to Theorem 4.6, that any *quasi-idempotent can* be split, assuming nothing more than function extensionality. There is an obvious naive thing to try: just as  $J$  extends  $I$  with an additional coherence, we might try to extend the type  $\sum_{x:X}(f(x) = x)$  that worked sometimes in §3 with an additional coherence, defining

$$\sum_{x:X} \sum_{p:f(x)=x} (\text{ap}_f(p) = I(x)). \quad (5.1)$$

However, Kraus has shown that this does not work in general.

**Example 5.1** (Kraus). Let  $X$  be a type with an element  $x_0 : X$ , and define  $f : X \rightarrow X$  by  $f(x) := x_0$  for all  $x : X$ . Then  $f$  is quasi-idempotent with  $I(x) := \text{refl}_{x_0}$  and  $J(x) := \text{refl}_{\text{refl}_{x_0}}$  for all  $x$ . However, the type of (5.1) in this case becomes

$$\sum_{x:X} \sum_{p:x_0=x} (\text{refl}_{x_0} = \text{refl}_{x_0}).$$

This is equivalent to

$$\sum_{q:\sum_{x:X}(x_0=x)} (\text{refl}_{x_0} = \text{refl}_{x_0})$$

and thence to simply  $\text{refl}_{x_0} = \text{refl}_{x_0}$ , since the type  $\sum_{x:X}(x_0 = x)$  is contractible.

On the other hand,  $f$  has an evident splitting with  $A := \mathbf{1}$ , where  $s$  picks out the point  $x_0$ . Thus, if (5.1) were also a splitting of it, then by Lemma 3.4 it would be equivalent to  $\mathbf{1}$ , i.e. contractible.

However, assuming univalence, there are pointed types  $(X, x_0)$  for which  $\text{refl}_{x_0} = \text{refl}_{x_0}$  is not contractible. For instance, we can take  $X$  to be the universe  $\mathbf{Type}$ , with  $x_0 := B\text{Aut}(\mathbf{2})$ . In this case, a nontrivial element of  $\text{refl}_{x_0} = \text{refl}_{x_0}$  is constructed in the proof of [Uni13, Theorem 4.1.3] (we will give a slightly different construction of the same element in Remark 8.4). Thus, MLTT cannot prove that (5.1) always splits a quasi-idempotent.

Thus thwarted in our naïve attempts, we turn again to  $\infty$ -category theory. The proof in [Lur14, Lemma 7.3.5.14] shows that one extra coherence datum suffices to construct a splitting as the colimit of the infinite sequence

$$X \xrightarrow{f} X \xrightarrow{f} X \xrightarrow{f} \dots$$

As observed before, colimits are difficult to handle in type theory. Fortunately, idempotents are completely self-dual, so we might just as well consider the limit of the infinite sequence

$$\dots \xrightarrow{f} X \xrightarrow{f} X \xrightarrow{f} X.$$

This is easy to define in type theory: it is  $\sum_{a:\mathbb{N} \rightarrow X} \prod_{n:\mathbb{N}} (f(a_{n+1}) = a_n)$ . Here we see the need for function extensionality: this type involves functions, and we need to construct equalities in it to exhibit it as a retract of  $X$ . For reference, we record exactly how to construct equalities in this type.

**Lemma 5.2.** *Given  $(a, \alpha)$  and  $(b, \beta)$  in  $\sum_{a:\mathbb{N} \rightarrow X} \prod_{n:\mathbb{N}} (f(a_{n+1}) = a_n)$ , to show that they are equal (assuming function extensionality) it is necessary and sufficient to*

- (1) *Construct for each  $n : \mathbb{N}$  an equality  $\xi_n : a_n = b_n$ , and*
- (2) *Show that for each  $n : \mathbb{N}$  the following diagram of equalities commutes:*

$$\begin{array}{ccc} f(a_{n+1}) & \xrightarrow{\alpha_n} & a_n \\ \text{ap}_f(\xi_{n+1}) \downarrow & & \downarrow \xi_n \\ f(b_{n+1}) & \xrightarrow{\beta_n} & b_n, \end{array} \quad (5.2)$$

*i.e. that  $\alpha_n \cdot \xi_n = \text{ap}_f(\xi_{n+1}) \cdot \beta_n$ .*

*Proof.* A straightforward application of the results of [Uni13, Chapter 2].  $\square$

Now we can prove the main theorem of this section.

**Theorem 5.3.** *Assuming function extensionality, any quasi-idempotent splits.*

*Proof.* Given  $(f, I, J)$ , define  $A := \sum_{a:\mathbb{N} \rightarrow X} \prod_{n:\mathbb{N}} (f(a_{n+1}) = a_n)$  as above. We define  $s : A \rightarrow X$  by  $s(a, \alpha) := a_0$ , and  $r : X \rightarrow A$  by the slightly less obvious formula

$$r(x) := (\lambda n. f(x), \lambda n. I(x)).$$

(Note that both components of  $r(x)$  are actually constant functions, i.e. independent of  $n$ .) Now we obviously have  $s \circ r = f$ ; the tricky part is proving  $r \circ s = 1$ .

Let  $(a, \alpha) : A$ , so that  $a : \mathbb{N} \rightarrow X$  and  $\alpha : \prod_{n:\mathbb{N}} (f(a_{n+1}) = a_n)$ . We must show  $(a, \alpha) = r(s(a, \alpha))$ . By definition, both components of  $r(s(a, \alpha))$  are constant, the first at  $f(a_0)$  and the second at  $I(a_0)$ ; thus we need a family of equalities  $a_n = f(a_0)$  that satisfy commutativity relations. For convenience, we break this down into two steps, by defining an intermediate element  $(b, \beta) : A$  and showing that  $(b, \beta) = (a, \alpha)$  and also  $(b, \beta) = r(s(a, \alpha))$ . The definition is

$$\begin{aligned} b_n &::= f(f(a_{n+1})) && : X \\ \beta_n &::= \text{ap}_{f \circ f}(\alpha_{n+1}) && : f(f(f(a_{n+2}))) = f(f(a_{n+1})) \end{aligned}$$

To show that  $(b, \beta) = (a, \alpha)$ , we apply Lemma 5.2 with

$$\xi_n ::= I(a_{n+1}) \cdot \alpha_n \quad : b_n = a_n.$$

We thus have to show that

$$\text{ap}_{f \circ f}(\alpha_{n+1}) \cdot I(a_{n+1}) \cdot \alpha_n = \text{ap}_f(I(a_{n+2}) \cdot \alpha_{n+1}) \cdot \alpha_n$$

which (after cancelling  $\alpha_n$ ) we can do as follows:

$$\begin{aligned} \text{ap}_{f \circ f}(\alpha_{n+1}) \cdot I(a_{n+1}) &= I(f(a_{n+2})) \cdot \text{ap}_f(\alpha_{n+1}) && \text{(naturality)} \\ &= \text{ap}_f(I(a_{n+2})) \cdot \text{ap}_f(\alpha_{n+1}) && \text{(by } J(a_{n+2})) \\ &= \text{ap}_f(I(a_{n+2}) \cdot \alpha_{n+1}) && \text{(functoriality)} \end{aligned}$$

Next we have to show that  $(b, \beta) = r(s(a, \alpha))$ . By definition,  $r(s(a, \alpha)) ::= (\lambda n. f(a_0), \lambda n. I(a_0))$ . Invoking Lemma 5.2 again, we need to firstly construct  $\xi_n : f(f(a_{n+1})) = f(a_0)$  for all  $n$ . We do this by induction on  $n$ . The base case  $n \equiv 0$  is simply  $\text{ap}_f(\alpha_0) : f(f(a_1)) = f(a_0)$ , while the induction step is the composite

$$f(f(a_{n+2})) = f(a_{n+1}) = f(f(a_{n+1})) = f(a_0)$$

of  $\text{ap}_f(\alpha_{n+1})$  and  $I(a_{n+1})^{-1}$  with the induction hypothesis.

It remains to show that  $\mathbf{ap}_{f \circ f}(\alpha_{n+1}) \cdot \xi_n = \mathbf{ap}_f(\xi_{n+1}) \cdot I(a_0)$  for all  $n$ , and we do this by induction on  $n$  as well. For the base case  $n \equiv 0$ , this means to check that

$$\mathbf{ap}_{f \circ f}(\alpha_1) \cdot \mathbf{ap}_f(\alpha_0) = \mathbf{ap}_f(\mathbf{ap}_f(\alpha_1) \cdot I(a_1)^{-1} \cdot \mathbf{ap}_f(\alpha_0)) \cdot I(a_0).$$

Applying functoriality of  $\mathbf{ap}_f$  on the right, canceling a copy of  $\mathbf{ap}_{f \circ f}(\alpha_1)$  on both sides, and rearranging a little this becomes

$$\mathbf{ap}_f(I(a_1)) \cdot \mathbf{ap}_f(\alpha_0) = \mathbf{ap}_{f \circ f}(\alpha_0) \cdot I(a_0).$$

But using  $J$  we can make this into

$$I(f(a_1)) \cdot \mathbf{ap}_f(\alpha_0) = \mathbf{ap}_{f \circ f}(\alpha_0) \cdot I(a_0).$$

which is an instance of naturality for  $I$ .

Finally, for the induction step, our inductive hypothesis is that the following diagram commutes:

$$\begin{array}{ccc} f(f(f(a_{n+2}))) & \xrightarrow{\mathbf{ap}_f(\mathbf{ap}_f(\alpha_{n+1}) \cdot I(a_{n+1})^{-1} \cdot \xi_n)} & f(f(a_0)) \\ \mathbf{ap}_{f \circ f}(\alpha_{n+1}) \downarrow & & \downarrow I(a_0) \\ f(f(a_{n+1})) & \xrightarrow{\xi_n} & f(a_0) \end{array}$$

and our goal is (after applying functoriality of  $\mathbf{ap}_f$ ) to prove that the outer boundary of the following diagram commutes.

$$\begin{array}{ccccccc} f(f(f(a_{n+3}))) & \xrightarrow{\mathbf{ap}_{f \circ f}(\alpha_{n+2})} & f(f(a_{n+2})) & \xrightarrow[\text{(J)}]{\mathbf{ap}_f(I(a_{n+2}))^{-1}} & f(f(f(a_{n+2}))) & \longrightarrow & f(f(a_0)) \\ \mathbf{ap}_{f \circ f}(\alpha_{n+2}) \downarrow & \nearrow \text{refl} & & \text{(nat)} \quad I(f(a_{n+2}))^{-1} & \downarrow & \text{(IH)} & \downarrow I(a_0) \\ f(f(a_{n+2})) & \xrightarrow{\mathbf{ap}_f(\alpha_{n+1})} & f(a_{n+1}) & \xrightarrow{I(a_{n+1})^{-1}} & f(f(a_{n+1})) & \xrightarrow{\xi_n} & f(a_0) \end{array}$$

The square marked (IH) is just the inductive hypothesis, while what remains can be filled in by naturality and a further application of  $J$ .  $\square$

**Remark 5.4.** Note that the type  $A$  and the section  $s : A \rightarrow X$  involved in the splitting can be defined without knowing either  $I$  or  $J$ , while the retraction  $r : X \rightarrow A$  and the homotopy  $K : s \circ r \sim f$  require only  $I$ . It is only the other homotopy  $H : r \circ s \sim \text{id}_A$  that requires the extra coherence datum  $J$ , and likewise only this homotopy that requires function extensionality.

## 6. SPLITTING IS A RETRACTION

**Assumption 6.1.** *In this section we assume the univalence axiom.*

Recall from the introduction that we can define the types of retractions of  $X$  and of quasi-idempotents on  $X$ :

$$\begin{aligned} \text{Retr}(X) &::= \sum_{A:\text{Type}} \sum_{r:X \rightarrow A} \sum_{s:A \rightarrow X} \prod_{a:A} (r(s(a)) = a) \\ \text{Qldem}(X) &::= \sum_{f:X \rightarrow X} \sum_{I:f \circ f \sim f} \prod_{x:X} (\mathbf{ap}_f(I(x)) = I(f(x))) \end{aligned}$$

Now Theorem 5.3 defines a map

$$\text{split} : \text{QIdem}(X) \rightarrow \text{Retr}(X)$$

and Lemma 3.6 defines a map

$$\text{uli} : \text{Retr}(X) \rightarrow \text{QIdem}(X).$$

We will now prove the following theorem.

**Theorem 6.2.** *split and uli exhibit  $\text{Retr}(X)$  as a retract of  $\text{QIdem}(X)$ . In other words,  $\text{split} \circ \text{uli} = \text{id}_{\text{Retr}(X)}$ .*

Before proving this, however, we need to know how to construct equalities in  $\text{Retr}(X)$ .

**Lemma 6.3.** *Suppose given  $(A, r, s, H)$  and  $(A', r', s', H')$  in  $\text{Retr}(X)$ . To give an equality  $(A, r, s, H) = (A', r', s', H')$ , it suffices to give*

- (1) *An equivalence  $A \simeq A'$ , including functions  $g : A \rightarrow A'$  and  $h : A' \rightarrow A$  and homotopies  $\eta : gh \sim \text{id}$  and  $\epsilon : hg \sim \text{id}$ ;*
- (2) *A homotopy  $P : hr' \sim r$ ;*
- (3) *A homotopy  $Q : s'g \sim s$ ; and*
- (4) *A witness that  $\text{ap}_h(H'(g(a))) \cdot \epsilon_a = P(s'(g(a))) \cdot \text{ap}_r(Q(a)) \cdot H(a)$  for all  $a : A$ .*

*Proof.* This should be considered a straightforward application of the results of [Uni13, Chapter 2] characterizing identity types of types obtained from different type-formers. Since  $\text{Retr}(X)$  is a triple  $\Sigma$ -type, by [Uni13, Theorem 2.7.2] we can first decompose equalities in  $\text{Retr}(X)$  as quadruples of equalities in its constituent types. We then apply univalence to obtain an equivalence  $A \simeq A'$  as the first component, [Uni13, Lemma 2.9.6] to obtain  $P$  and  $Q$  as the second and third, and similarly for the fourth component. The details are tedious, so we leave them to the reader; like the rest of the paper they have been formalized in Coq.  $\square$

*Proof of Theorem 6.2.* Suppose given a retraction  $(A, r : X \rightarrow A, s : A \rightarrow X, H : r \circ s \sim 1)$ ; we want to show that it is equivalent to the splitting of the induced quasi-idempotent  $sr : X \rightarrow X$ . The latter is a new retraction  $(A', r', s', H')$  such that  $sr = s'r'$  (in fact this equality holds judgmentally). By Lemma 3.4, we have an equivalence  $A \simeq A'$  composed of  $g = r's : A \rightarrow A'$  and  $h = rs' : A' \rightarrow A$ , with  $\eta : gh \equiv r'srs' = r's'r's' = \text{id}$  and dually  $\epsilon : hg \equiv rs'r's = rsrs = \text{id}$ . Moreover, we have  $P : hr' \equiv rs'r' = rsr = r$  and  $Q : s'g \equiv s'r's = srs = s$ ; thus it remains to construct the fourth datum in Lemma 6.3.

In general, both sides of this equality are homotopies  $rs'r's'r's \sim 1$ . Note that in our case, the domain  $rs'r's'r's$  is judgmentally equal to  $rsrsrs$ . Substituting the definitions of  $r'$ ,  $s'$ , and  $H'$  from Theorem 5.3, we see that the left-hand side of the desired equality is the composite

$$\begin{aligned} rsrsrsa &\xrightarrow{\text{ap}_{rsrs}(H(rsa))^{-1}} rsrsrsrsa \xrightarrow{\text{ap}_{rs}(H(rsrsa))} rsrsrsa \\ &\xrightarrow{\text{ap}_{rs}(H(rsa))} rsrsa \xrightarrow{H(rsa)} rsa \xrightarrow{Ha} a \end{aligned} \quad (6.1)$$

while the right-hand side is the composite

$$rsrsrsa \xrightarrow{H(rsrsa)} rsrsa \xrightarrow{\text{ap}_{rs}(Ha)} rsa \xrightarrow{Ha} a. \quad (6.2)$$

Now by naturality, we have

$$\text{ap}_{rs}(H(rsrsa)) \cdot \text{ap}_{rs}(H(rsa)) = \text{ap}_{rsrs}(H(rsa)) \cdot \text{ap}_{rs}(H(rsa)).$$

Applying this in the middle of (6.1), and canceling  $\mathbf{ap}_{rs}(H(rsa))$  with its inverse on the left, we reduce it to

$$rsrsrsa \xrightarrow{\mathbf{ap}_{rs}(H(rsa))} rsrsra \xrightarrow{H(rsa)} rsa \xrightarrow{Ha} a.$$

Now naturality gives  $\mathbf{ap}_{rs}(H(rsa)) \cdot H(rsa) = H(rsrsa) \cdot H(rsa)$ , so this is equal to

$$rsrsrsa \xrightarrow{H(rsrsa)} rsrsra \xrightarrow{H(rsa)} rsa \xrightarrow{Ha} a.$$

Comparing this to (6.2), we can cancel  $H(rsrsa)$  on the left, reducing the problem to  $\mathbf{ap}_{rs}(Ha) \cdot Ha = H(rsa) \cdot Ha$ , which is another naturality.  $\square$

Theorem 6.2 makes no reference to a specified function  $f$ , but we can deduce from it a statement that does. Given an endofunction  $f$ , we define a **splitting of  $f$**  to be a retraction  $(A, r, s, H) : \mathbf{Retr}(X)$  together with a homotopy  $K : s \circ r \sim f$ . These form a type  $\mathbf{Split}(X, f) := \sum_{(A, r, s, H) : \mathbf{Retr}(X)} (s \circ r \sim f)$ . We also have a type  $\mathbf{QIdem}(X, f) := \sum_{(I : f \circ f \sim f)} \prod_{x : X} (\mathbf{ap}_f(I(x)) = I(f(x)))$  of “quasi-idempotence data for  $f$ ”.

**Corollary 6.4.** *For any  $f : X \rightarrow X$ , the type  $\mathbf{Split}(X, f)$  is a retract of  $\mathbf{QIdem}(X, f)$ .*

*Proof.* Let  $k : \mathbf{Retr}(X) \rightarrow (X \rightarrow X)$  take  $(A, r, s, H)$  to the composite  $sr$ . Then  $\mathbf{Split}(X, f)$  is, by definition, the fiber of  $k$  over  $f$ . On the other hand, by [Uni13, Lemma 4.8.1], the type  $\mathbf{QIdem}(X, f)$  is equivalent to the fiber over  $f$  of the first projection  $\mathbf{QIdem}(X) \rightarrow (X \rightarrow X)$ . The retraction from Theorem 6.2 commutes with these maps to  $X \rightarrow X$ ; hence by [Uni13, Lemma 4.7.3], it induces a retraction between their fibers.  $\square$

Similarly, we can consider the case when  $f$  is already equipped with a witness  $I$  of pre-idempotency. We define  $\mathbf{Split}(X, f, I)$  to be

$$\sum_{(A, r, s, H, K) : \mathbf{Split}(X, f)} \prod_{x : X} (\mathbf{ap}_f(K(x))^{-1} \cdot K(s(r(x)))^{-1} \cdot \mathbf{ap}_s(H(r(x))) \cdot K(x) = I(x)),$$

the long composite being just the result of transferring the definition of Lemma 3.3 across the homotopy  $K : sr \sim f$ . And of course we have the type  $\mathbf{QIdem}(X, f, I) := \prod_{x : X} (\mathbf{ap}_f(I(x)) = I(f(x)))$  of quasi-idempotence enhancements of  $I$ .

**Corollary 6.5.** *For any  $(f, I)$ , the type  $\mathbf{Split}(X, f, I)$  is a retract of  $\mathbf{QIdem}(X, f, I)$ .*

*Proof.* As in Corollary 6.4, we take fibers of two maps to the type  $f \circ f \sim f$  of  $I$ . We leave the details to the reader; or they can be found in the formalization.  $\square$

## 7. SPLITTING IS NOT AN EQUIVALENCE

We now consider what can be said about the composite  $\mathbf{uli} \circ \mathbf{split}$ , which is an endofunction of  $\mathbf{QIdem}(X)$ . Our first observation is that it preserves the witness  $I$  of pre-idempotence.

**Theorem 7.1.** *Assume function extensionality. Then given a quasi-idempotent  $(f, I, J)$ , if we split it as in Theorem 5.3, the witness of pre-idempotence induced from the splitting as in Lemma 3.3 is equal to  $I$ .*

*Proof.* Given a retraction  $s : A \rightarrow X$  and  $r : X \rightarrow A$  with  $H : r \circ s \sim 1$ , the induced  $I$  was defined in Lemma 3.3 by  $I(x) := \mathbf{ap}_s(H(r(x)))$ . For the splitting from Theorem 5.3 with  $A := \sum_{a : \mathbb{N} \rightarrow X} \prod_{x : X} (f(a_{n+1}) = a_n)$ , we have  $s(a, b) := a_0$ , so the induced  $I'(x)$  is just the 0-component of the homotopy  $H : r \circ s \sim 1$  at  $r(x) := (\lambda n. f(x), \lambda n. I(x))$ . By construction,



this is the composite

$$f(f(x)) \xrightarrow{\mathbf{ap}_f(I(x))^{-1}} f(f(f(x))) \xrightarrow{I(f(x))} f(f(x)) \xrightarrow{I(x)} f(x)$$

where  $I$  is the given witness of pre-idempotence. But by the given  $J$ , we have  $\mathbf{ap}_f(I(x)) = I(f(x))$ , so this reduces to just  $I(x)$ .  $\square$

Thus, if the further coherence witness  $J$  were also recovered from the splitting, we would have  $\mathbf{uli} \circ \mathbf{split} = \mathbf{id}$ , and hence (assuming univalence, so that the results of the previous section apply)  $\mathbf{split}$  and  $\mathbf{uli}$  would be inverse *equivalences* between  $\mathbf{Retr}(X)$  and  $\mathbf{Qldem}(X)$ . By Corollary 6.4 and Corollary 6.5, this would also yield equivalences between  $\mathbf{Split}(X, f)$  and  $\mathbf{Qldem}(X, f)$ , and between  $\mathbf{Split}(X, f, I)$  and  $\mathbf{Qldem}(X, f, I)$ , for any  $f$  and  $I$ . We will show that this is impossible in general, beginning with the following observation.

**Lemma 7.2.** *Assuming univalence, if  $f \equiv \mathbf{id}_X$  and  $I(x) \equiv \mathbf{refl}_x$  for all  $x$ , then the type  $\mathbf{Split}(X, f, I)$  is contractible.*

*Proof.* Recall that for any type  $B$  and point  $b_0 : B$ , the type  $\sum_{b:B} (b = b_0)$  is contractible. By univalence, it follows that for any type  $X$ , the type  $\sum_{A:\mathbf{Type}} (A \simeq X)$  is contractible. Since  $\mathbf{Split}(X, f, I)$  begins with a  $\sum_{A:\mathbf{Type}}$ , it will suffice to show that the rest of it is equivalent to  $(A \simeq X)$ .

We will use the “half-adjoint equivalence” definition of  $(A \simeq X)$  from [Uni13, §4.2]. The data  $r$  and  $s$  are, of course, maps back and forth, while since  $f \equiv \mathbf{id}_X$  the data  $H$  and  $K$  have the right types to be the homotopies  $\epsilon$  and  $\eta$ . It remains, therefore, to show that the type of the remaining datum:

$$\prod_{x:X} (\mathbf{ap}_f(K(x))^{-1} \cdot K(s(r(x)))^{-1} \cdot \mathbf{ap}_s(H(r(x))) \cdot K(x) = I(x))$$

is equivalent to  $\prod_{a:A} (\mathbf{ap}_s(H(a)) = K(s(a)))$ . Now since  $f \equiv \mathbf{id}_X$  and  $I(x) \equiv \mathbf{refl}_x$ , we can discard the  $\mathbf{ap}_f$ , move the  $K(x)^{-1}$  to the other side, and then cancel it. If we move  $K(s(r(x)))$  to the other side as well, we obtain  $\prod_{x:X} (\mathbf{ap}_s(H(r(x))) = K(s(r(x))))$ . Finally, since  $s$ ,  $H$ , and  $K$  suffice to show that  $r$  is an equivalence, we can transport along it to obtain the desired type  $\prod_{a:A} (\mathbf{ap}_s(H(a)) = K(s(a)))$ .  $\square$

Therefore, if we had  $\mathbf{uli} \circ \mathbf{split} = \mathbf{id}$ , then  $\mathbf{Qldem}(X, \mathbf{id}_X, \lambda x. \mathbf{refl}_x)$  would also be contractible for any  $X$ . However,  $\mathbf{Qldem}(X, \mathbf{id}_X, \lambda x. \mathbf{refl}_x)$  reduces to  $\prod_{x:X} (\mathbf{refl}_x = \mathbf{refl}_x)$ , which we might call the **2-center** of  $X$  (see §8 for why). Thus, it suffices to construct a type  $X$  whose 2-center has nontrivial inhabitants. Of course, such an  $X$  cannot be a set or even a 1-type, but it will suffice for it to be a 2-type (i.e. its twice-iterated equality types  $p =_{(x=Xy)} q$  are sets).

**Remark 7.3.** As pointed out by a referee, there are many ways to construct such a 2-type using higher inductive types. For instance, if  $X$  is the 2-truncation of the 2-sphere, we can define an element of  $\prod_{x:X} (\mathbf{refl}_x = \mathbf{refl}_x)$  by truncation-induction (since  $\mathbf{refl}_x = \mathbf{refl}_x$  is a 0-type) followed by sphere-induction, sending the basepoint to the generating 2-loop and the rest being trivial for truncation reasons. More generally, we could take  $X$  to be an Eilenberg–Mac Lane space  $K(G, 2)$  for any nontrivial abelian group  $G$  (see [LF14]) — the 2-truncation of the 2-sphere is a  $K(\mathbb{Z}, 2)$ . However, if we are willing to work a little harder, we can obtain such a 2-type using only univalence and propositional truncation: just as  $\mathbf{BAut}(\mathbf{2})$  supports a nontrivial element of the **1-center**  $\prod_{x:X} (x = x)$ , to find a nontrivial element of the 2-center we can use  $X \equiv \mathbf{BAut}(\mathbf{BAut}(\mathbf{2}))$ .

**Theorem 7.4.** *Assuming univalence and propositional truncation, if  $X := \mathbf{BAut}(\mathbf{BAut}(\mathbf{2}))$ , then  $\prod_{x:X}(\mathbf{refl}_x = \mathbf{refl}_x)$  has a nontrivial element.*

*Idea of proof.* As an  $\infty$ -groupoid,  $\mathbf{BAut}(\mathbf{2})$  has one object with two automorphisms, the identity and the flip. Since automorphisms preserve identities,  $\mathbf{BAut}(\mathbf{2})$  itself has only one automorphism, but there are two self-homotopies of that automorphism. In other words, the space of automorphisms of  $\mathbf{BAut}(\mathbf{2})$  is equivalent to  $\mathbf{BAut}(\mathbf{2})$  itself. Thus,  $\mathbf{BAut}(\mathbf{BAut}(\mathbf{2}))$  has one object, with only its identity morphism, but two 2-morphisms from that identity to itself. This nonidentity 2-morphism is essentially our desired nontrivial element.

However, proving this carefully in type theory requires a lot of lemmas about classifying spaces, so we defer it to the next section. (An alternative proof can be found in [Kra15, Lemma 7.5.2].)  $\square$

**Corollary 7.5.** *In MLTT with function extensionality (which is necessary to construct the function  $\mathbf{split}$ ), it is impossible to prove that  $\mathbf{uli} \circ \mathbf{split} = \mathbf{id}_{\mathbf{QIdem}(X)}$  for every type  $X$ .*  $\square$

## 8. THE DOUBLE CLASSIFYING SPACE OF 2

Here we will prove Theorem 7.4. For this we need some preliminary lemmas about types of the form  $\mathbf{BAut}(X)$ .

**Assumption 8.1.** *Throughout this section we assume both univalence and propositional truncation.*

Our first lemma says that defining a section of a family of sets indexed by  $\mathbf{BAut}(X)$  is equivalent to giving an element lying over  $X$  itself which is fixed by all automorphisms of  $X$ . To make sense of “fixed by”, we use the notion of **transport**: given any type family  $B : A \rightarrow \mathbf{Type}$ , if  $p : x =_A y$  we have a function  $p_* : B(x) \rightarrow B(y)$  defined by identity-type elimination (see [Uni13, Chapter 2] for more information).

For convenience, we will frequently implicitly coerce elements of  $\mathbf{BAut}(X)$  to their underlying types.

**Lemma 8.2.** *Let  $X$  be any type, and suppose  $P : \mathbf{BAut}(X) \rightarrow \mathbf{Type}$  is a family of sets. Then*

$$\left( \prod_{Z:\mathbf{BAut}(X)} P(Z) \right) \simeq \left( \sum_{e:P(X)} \prod_{g:X=X} g_*(e) = e \right).$$

*Proof.* Since  $\mathbf{BAut}(X) := \sum_{Z:\mathbf{Type}} \|Z = X\|$ , we have

$$\left( \prod_{Z:\mathbf{BAut}(X)} P(Z) \right) \simeq \left( \prod_{Z:\mathbf{Type}} (\|Z = X\| \rightarrow P(Z)) \right). \quad (8.1)$$

Now recall from [KECA14] that if  $B$  is a set, then a function  $A \rightarrow B$  factors through  $\|A\|$  if and only if it is weakly constant. In fact, it is not hard to show that when  $B$  is a set, the type  $\|A\| \rightarrow B$  is equivalent to the type of weakly constant functions  $A \rightarrow B$ . Thus, the right-hand-side of (8.1) is equivalent to

$$\prod_{Z:\mathbf{Type}} \sum_{f:(Z=X) \rightarrow P(Z)} \prod_{p,q:Z=X} (f(p) = f(q)).$$

Rearranging this with [Uni13, Theorem 2.15.7] (the “type-theoretic axiom of choice”), we obtain

$$\sum_{f:\prod_{Z:\mathbf{Type}} (Z=X) \rightarrow P(Z)} \prod_{Z:\mathbf{Type}} \prod_{p,q:Z=X} (f(Z,p) = f(Z,q)).$$

Applying the universal property of identity types [Uni13, (2.15.10)], this becomes

$$\sum_{f:\prod_{Z:\mathbf{Type}}(Z=X)\rightarrow P(Z)} \prod_{p:X=X} (f(X,p) = f(X, \mathbf{refl}_X)).$$

The same property implies that  $\prod_{Z:\mathbf{Type}}(Z=X) \rightarrow P(Z)$  is equivalent to  $P(X)$ , where the inverse equivalence sends  $e : P(X)$  to  $\lambda Z.\lambda q.q_*(e)$ . Transferring across this equivalence, we obtain the desired result.  $\square$

We can use this to characterize types of the form  $\prod_{Z:BAut(X)}(Z=Z)$ , which are “one level down” from the type  $\prod_{Z:BAut(X)}(\mathbf{refl}_Z = \mathbf{refl}_Z)$  considered in Theorem 7.4.

**Lemma 8.3.** *If  $X$  is a set, then  $\prod_{Z:BAut(X)}(Z=Z)$  is equivalent to*

$$\sum_{f:X\simeq X} \prod_{g:X\simeq X} (f \circ g = g \circ f)$$

*Proof.* Since  $X$  is a set,  $Z=Z$  is a set for any  $Z : BAut(X)$ . Thus, by Lemma 8.2,  $\prod_{Z:BAut(X)}(Z=Z)$  is equivalent to

$$\sum_{e:X=X} \prod_{g:X=X} (g_*(e) = e).$$

The result follows by applying [Uni13, Theorem 2.11.5] and the univalence axiom.  $\square$

**Remark 8.4.** If  $X := \mathbf{2}$ , it is easy to show that  $X$  has precisely two automorphisms, the identity and the flip. Since the flip is an involution, it commutes with itself, and of course it commutes with the identity; thus by Lemma 8.3 it yields a nontrivial element of  $\prod_{Z:BAut(\mathbf{2})}(Z=Z)$ . This gives a slightly different proof of [Uni13, Theorem 4.1.3]. In fact, Lemma 8.3 gives the stronger result that  $\prod_{Z:BAut(\mathbf{2})}(Z=Z)$  has *exactly* one nontrivial element (hence in particular our nontrivial element agrees with that of [Uni13, Theorem 4.1.3]).

Lemma 8.3 says that  $\prod_{Z:BAut(X)}(Z=Z)$  is equivalent to the type of automorphisms of  $X$  that commute with all other automorphisms of  $X$ , i.e. the **center** of  $Aut(X)$ . This explains why when we move up a level to the type appearing in Theorem 7.4, we may reasonably call it the **2-center**.

**Lemma 8.5.** *If  $X$  is a 1-type, then  $\prod_{Z:BAut(X)}(\mathbf{refl}_Z = \mathbf{refl}_Z)$  is equivalent to*

$$\sum_{f:\prod_{x:X}(x=x)} \prod_{g:X\simeq X} \prod_{x:X} (\mathbf{ap}_g(f(x)) = f(g(x))).$$

*Proof.* Since  $X$  is a 1-type,  $(\mathbf{refl}_Z = \mathbf{refl}_Z)$  is a set for any  $Z : BAut(X)$ . Thus, by Lemma 8.2,  $\prod_{Z:BAut(X)}(\mathbf{refl}_Z = \mathbf{refl}_Z)$  is equivalent to

$$\sum_{e:\mathbf{refl}_X = \mathbf{refl}_X} \prod_{g:X=X} g_*(e) = e.$$

Now by univalence and function extensionality,  $\mathbf{refl}_X = \mathbf{refl}_X$  is equivalent to  $\prod_{x:X}(x=x)$ , while of course  $X=X$  is equivalent to  $X\simeq X$ . Under this equivalence,  $g_*(e)$  is identified with  $\lambda x.\mathbf{ap}_g(f(g^{-1}(x)))$ . Finally, since  $g$  is an equivalence, we can transfer it to the other side of the equation and obtain the desired result.  $\square$

We want to apply Lemma 8.5 to  $X := BAut(\mathbf{2})$ . In that case, we have a nontrivial  $f : \prod_{x:BAut(\mathbf{2})}(x=x)$  from Remark 8.4. Therefore, to prove Theorem 7.4 it remains to show that this  $f$  satisfies  $\mathbf{ap}_g(f(Z)) = f(g(Z))$  for all automorphisms  $g$  of  $BAut(\mathbf{2})$  and all  $Z : BAut(\mathbf{2})$ .

Of course, this requires knowing something about *all* automorphisms of  $BAut(\mathbf{2})$ . In our proof sketch of Theorem 7.4, we claimed that the space of automorphisms of  $BAut(\mathbf{2})$  should be equivalent to  $BAut(\mathbf{2})$  itself, but our argument involved decomposing an  $\infty$ -groupoid into “objects, morphisms, and 2-morphisms” which is not possible in homotopy type theory. Instead, we need to give a more “synthetic” argument, analogous to our construction of the incoherent pre-idempotent on  $BAut(C)$  in §4.

The idea is as follows: since  $\mathbf{2}$  is an abelian group (the cyclic group of order 2),  $BAut(\mathbf{2})$  should also be an abelian  $\infty$ -group. Since multiplication by a fixed element of an  $\infty$ -group is an equivalence, this will give us a map  $BAut(\mathbf{2}) \rightarrow (BAut(\mathbf{2}) \simeq BAut(\mathbf{2}))$ , which we can then show to be an equivalence.

Now we have to define the group operation on  $BAut(\mathbf{2})$  internally. The idea to keep in mind is that the elements of  $BAut(\mathbf{2})$  are the “finite sets with two elements”. They are merely isomorphic to  $\mathbf{2}$ , but to *specify* such an isomorphism  $\mathbf{2} \simeq Z$  is the same as specifying an element of  $Z$  (to be the image of  $1 : \mathbf{2}$ ).

The “morally-best” definition of the group operation would perhaps be as a “tensor product over the field with two elements”. However, since we are not assuming any colimits, we use instead the following:

$$Z * W := (Z \simeq W).$$

Since  $\mathbf{2} \simeq (\mathbf{2} \simeq \mathbf{2})$ , it follows that  $Z \simeq W$  is in  $BAut(\mathbf{2})$  if  $Z$  and  $W$  are.

This definition is obviously symmetric,  $Z * W = W * Z$ . Moreover, it has  $\mathbf{2}$  itself as a left (hence also right) identity: if  $W : BAut(\mathbf{2})$  then an equivalence  $e : \mathbf{2} \simeq W$  is uniquely determined by  $e(1) : W$ . And  $Z * Z$  is equivalent to  $\mathbf{2}$  for any  $Z$ , since it has a canonically specified element (namely the identity); thus in particular  $*$  has inverses. The trickiest part is showing associativity.

**Lemma 8.6.** *For any  $Z, W, Y : BAut(\mathbf{2})$  we have  $(Z * W) * Y = Z * (W * Y)$ .*

*Proof.* Since  $*$  is symmetric, it suffices to prove  $Y * (Z * W) = Z * (Y * W)$ . We will show that for all  $Y, Z, W$  there is a map  $\sigma : Y * (Z * W) \rightarrow Z * (Y * W)$ , and that this map is its own inverse (when applied with  $Y$  and  $Z$  switched).

Now, an element of  $Y * (Z * W)$  can be regarded as a function  $e : Y \rightarrow (Z \rightarrow W)$  with the additional properties that

- (1) each function  $e(y) : Z \rightarrow W$  is an equivalence, and
- (2)  $e$  induces an equivalence from  $Y$  to  $Z \simeq W$ .

Since being an equivalence is a mere proposition, two elements of  $Y * (Z * W)$  are equal just when their underlying functions  $e : Y \rightarrow (Z \rightarrow W)$  are.

We will define  $\sigma$  so that its action on underlying functions simply swaps arguments:  $\sigma(e)(z)(y) = e(y)(z)$ . Thus, it will automatically be self-inverse. What remains is to show that  $\sigma(e)$  satisfies (1) and (2) assuming  $e$  does.

However, since all of our types are finite sets (that is, they are merely isomorphic to a standard finite type such as  $\sum_{k:\mathbb{N}} (k < n)$ ), a map between them is an equivalence as soon as it is injective. Thus, to show (1) for  $\sigma(e)$  we must show that if  $e(y)(z) = e(y')(z)$  for some  $z : Z$ , then  $y = y'$ . But by (2) for  $e$ , we have  $y = e^{-1}(e(y))$  and  $y' = e^{-1}(e(y'))$ , so it suffices to show that  $e(y) = e(y')$ . This follows from  $e(y)(z) = e(y')(z)$  since an equivalence between 2-element sets is determined by its action on a single element.

Similarly, to show (2) for  $\sigma(e)$ , we must show that if  $e(y)(z) = e(y)(z')$  for *all*  $y : Y$ , then  $z = z'$ . But this in particular implies that  $e(y)(z) = e(y)(z')$  for *some*  $y$ , and thus  $z = z'$  by (1) for  $e$ .  $\square$

Now we can prove that  $BAut(\mathbf{2})$  is equivalent to its own automorphism group.

**Lemma 8.7.**  $BAut(\mathbf{2}) \simeq (BAut(\mathbf{2}) \simeq BAut(\mathbf{2}))$ .

*Proof.* The map from left to right sends  $Z$  to  $\lambda W.Z * W$ . Since  $Z * (Z * W) = (Z * Z) * W = \mathbf{2} * W = W$ , the function  $\lambda W.Z * W$  is an equivalence whose inverse is itself.

The map from right to left sends  $e : BAut(\mathbf{2}) \simeq BAut(\mathbf{2})$  to  $e^{-1}(\mathbf{2})$ . The round-trip composite on the left is the identity since  $\mathbf{2}$  is a unit for  $*$ . On the other side, we must show that for any  $e : BAut(\mathbf{2}) \simeq BAut(\mathbf{2})$  and  $W$  we have  $e(W) = e^{-1}(\mathbf{2}) * W$ .

In fact, we will show that  $e^{-1}(Z) * W = Z * e(W)$  for any  $Z, W$ ; the desired result then follows by taking  $Z := \mathbf{2}$ . However, by univalence, we have  $(e^{-1}(Z) * W) = (e^{-1}(Z) = W)$  and similarly on the other side, and  $(e^{-1}(Z) = W) \simeq (Z = e(W))$  holds for any equivalence  $e$ .  $\square$

Finally, we can prove Theorem 7.4.

**Theorem 8.8.** *There is an element of  $\prod_{Z:BAut(BAut(\mathbf{2}))}(\text{refl}_Z = \text{refl}_Z)$  that is not equal to  $\lambda Z.\text{refl}_{\text{refl}_Z}$ .*

*Proof.* By Remark 8.4, we have an  $f : \prod_{x:BAut(\mathbf{2})}(x = x)$  that is unequal to  $\lambda x.\text{refl}_x$ . Thus, by Lemma 8.5, it remains to show that this  $f$  satisfies  $\text{ap}_g(f(Z)) = f(g(Z))$  for all automorphisms  $g$  of  $BAut(\mathbf{2})$  and all  $Z : BAut(\mathbf{2})$ .

Let  $g$  and  $Z$  be given. By Lemma 8.7, we may assume  $g$  is of the form  $\lambda Y.W * Y$  for some  $W : BAut(\mathbf{2})$ . And since our goal is a mere proposition, we may assume that  $Z$  and  $W$  are both  $\mathbf{2}$ .

Now since  $g(Y) \equiv \mathbf{2} * Y$  and  $\mathbf{2}$  is a left unit for  $*$ , we have a homotopy  $H : g \sim \text{id}$ . And by “dependent  $\text{ap}$ ” for  $f$  (see [Uni13, Lemma 2.3.4]) applied to  $H_{\mathbf{2}} : \mathbf{2} * \mathbf{2} = \mathbf{2}$ , we have

$$f(\mathbf{2} * \mathbf{2}) \cdot H_{\mathbf{2}} = H_{\mathbf{2}} \cdot f(\mathbf{2}).$$

Since also  $g(Z) \equiv \mathbf{2} * \mathbf{2}$ , what we have to show becomes

$$\text{ap}_g(f(\mathbf{2})) \cdot H_{\mathbf{2}} = H_{\mathbf{2}} \cdot f(\mathbf{2}).$$

However, this is just naturality for  $H$ .  $\square$

## 9. COHERENT IDEMPOTENTS

We have seen that, assuming univalence,  $\text{Retr}(X)$  is a retract of  $\text{QIdem}(X)$ , and in general a nontrivial one. As remarked in the introduction, in  $\infty$ -category theory the “space of retractions of  $X$ ” is equivalent to the “space of fully-coherent idempotents on  $X$ ”. This follows from [Lur09, Corollary 4.4.5.14]. As stated, that corollary says that in an  $\infty$ -category where (fully-coherent) idempotents split, the space of *all* retractions is equivalent to the space of *all* fully-coherent idempotents; but since this equivalence is fibered over the space of objects of the  $\infty$ -category itself, it induces fiberwise equivalences for each object  $X$ .

Thus, in homotopy type theory with the univalence axiom, it is reasonable to expect that  $\text{Retr}(X)$  should be equivalent to “the type of fully-coherent idempotents on  $X$ ”, if we were able to define the latter type. In particular, since  $\text{Retr}(X)$  is not generally equivalent to  $\text{QIdem}(X)$ , the latter is not a correct definition of the type of fully-coherent idempotents.

As mentioned in the introduction, we could take  $\text{Retr}(X)$  as a *definition* of the type of fully-coherent idempotents, but this would suffer from two drawbacks:

- (1) It would be aesthetically unsatisfying to say that “an idempotent” comes by definition equipped with a splitting. Morally, splitting should be something that is *done to* an idempotent.
- (2) It lives in a higher universe than the type  $X$ , since it involves a  $\sum_{A:\text{Type}}$ .

Both of these problems can be solved with the following observation: since  $\text{Retr}(X)$  is a retract of  $\mathbf{QIdem}(X)$ , the composite  $\text{uli} \circ \text{split}$  is a *quasi-idempotent* on  $\mathbf{QIdem}(X)$ . We can therefore *split it* using the construction of Theorem 5.3. By Lemma 3.4, the resulting type will be equivalent to  $\text{Retr}(X)$ ; but it will live (like  $\mathbf{QIdem}(X)$  itself) in the same universe as  $X$ , and its elements do not obviously contain a splitting. Thus, we propose the following definition.

**Definition 9.1.** A **(fully-coherent) idempotent** on a type  $X$  is an element of the splitting of  $\text{uli} \circ \text{split}$ . Somewhat more explicitly, this type is

$$\text{Idem}(X) := \sum_{a:\mathbb{N} \rightarrow \mathbf{QIdem}(X)} \prod_{n:\mathbb{N}} (\text{uli}(\text{split}(a_{n+1})) = a_n).$$

Similarly, an **idempotent structure on**  $f : X \rightarrow X$  is an element of the splitting of the similarly induced idempotent on  $\mathbf{QIdem}(X, f)$ .

It is worth thinking a little about what assumptions are necessary for this definition. It may appear at first to require univalence, since  $\text{uli} \circ \text{split}$  is only a (quasi-)idempotent because of Lemma 6.3, which uses univalence. However, as observed in Remark 5.4, to define the splitting *type* of an idempotent does not require the witnesses of quasi-idempotency or pre-idempotency. Thus, in order to define the type  $\text{Idem}(X)$  we really only require function extensionality, since that suffices to define the maps  $\text{uli}$  and  $\text{split}$ .

It is possible, of course, to unwind this definition further, but it becomes quite complicated. Nevertheless, it is satisfying that we can give *some* correct definition of fully-coherent idempotent, since the general problem of representing fully-coherent higher homotopy structures in type theory is unsolved.

There is an interesting analogy to the situation with equivalences. The naïve definition of an equivalence (or isomorphism) between types  $A$  and  $B$  would be

$$\sum_{f:A \rightarrow B} \sum_{g:B \rightarrow A} (g \circ f \sim \text{id}_A) \times (f \circ g \sim \text{id}_B). \quad (9.1)$$

However, this gives the wrong homotopy type. We might then think that we need an infinite tower of further coherences, but in fact it suffices to give one additional datum, although there are several choices for what that extra datum might be (see [Uni13, Chapter 4]).

Nevertheless, given an element of (9.1), it is possible to alter one of its constituent homotopies to obtain a fully-coherent equivalence. This exhibits the type of equivalences as a retract of (9.1), just as our type of idempotents is a retract of the type of quasi-idempotents. There is a difference, however, in that “ $f$  is an equivalence” is a mere proposition, whereas “ $f$  is an idempotent” is not.

## 10. CONCLUSIONS

The main result of this paper is that not all idempotents in Martin-Löf type theory can be proven to split, but if we assume function extensionality then one additional coherence condition suffices to make an idempotent splittable. In addition to its intrinsic interest, this shows how ideas from homotopy theory and higher category theory can be useful even for the study of non-homotopical type theory.

In the homotopical case, however, there is more to say about idempotents, which can be partially or fully coherent. Although fully coherent homotopical structures are often difficult to define in type theory, we have managed to define the type of fully coherent idempotents, by splitting an idempotent on the type of partially coherent ones.

With that said, this paper still leaves a number of interesting open questions about idempotents in type theory.

**Open Problem 10.1.** Can we split quasi-idempotents in MLTT without assuming function extensionality? In particular, is there any more “finite” way to construct such a splitting?

**Open Problem 10.2.** Is the section  $\text{Idem}(X) \rightarrow \text{QIdem}(X)$  an embedding? Equivalently, by Theorem 3.10, is the type  $\text{uli}(\text{split}(f, I, J)) = (f, I, J)$  collapsible for every quasi-idempotent  $(f, I, J)$ ? I expect the answer is no, but an explicit counterexample would be nice to have.

**Open Problem 10.3.** Similarly, is the induced map from  $\text{Idem}(X)$  to the type  $\text{PIdem}(X)$  of *pre*-idempotents an embedding? Again, I expect the answer is no, but this appears to be an open problem even in  $\infty$ -category theory; see [Shu14].

**Open Problem 10.4.** Can  $\text{Idem}(X)$  be defined without assuming even function extensionality? More precisely, is there a type we can define without function extensionality that becomes equivalent to  $\text{Idem}(X)$  if we assume function extensionality?

**Open Problem 10.5.** Are there any other fully-coherent higher-homotopy structures that can be obtained from a finite amount of coherence by splitting an idempotent?

#### ACKNOWLEDGEMENT

This paper would not exist without Martín Escardó: not just because he asked the original question and contributed many of the results in §3, but because during a long email discussion he provided both encouragement and an indispensable sounding-board for the development of the rest of it, and gave helpful feedback on a draft.

#### REFERENCES

- [APW13] Steve Awodey, Álvaro Pelayo, and Michael A. Warren. Voevodsky’s univalence axiom in homotopy type theory. *Notices Amer. Math. Soc.*, 60(9):1164–1167, 2013.
- [Awo12] Steve Awodey. Type theory and homotopy. In *Epistemology versus ontology*, volume 27 of *Log. Epistemol. Unity Sci.*, pages 183–201. Springer, Dordrecht, 2012.
- [HoT15] HoTT Project. The homotopy type theory Coq library. <http://github.com/HoTT/HoTT/>, 2015.
- [HS98] Martin Hofmann and Thomas Streicher. The groupoid interpretation of type theory. In *Twenty-five years of constructive type theory (Venice, 1995)*, volume 36 of *Oxford Logic Guides*, pages 83–111. Oxford Univ. Press, New York, 1998.
- [Joh02] Peter T. Johnstone. *Sketches of an Elephant: A Topos Theory Compendium: Volume 2*. Number 43 in Oxford Logic Guides. Oxford Science Publications, 2002.
- [KECA14] Nicolai Kraus, Martín Escardó, Thierry Coquand, and Thorsten Altenkirch. Notions of anonymous existence in Martin-Löf type theory. <http://www.cs.nott.ac.uk/~ngk/docs/existence.pdf>, 2014.
- [KLV12] Chris Kapulkin, Peter LeFanu Lumsdaine, and Vladimir Voevodsky. The simplicial model of univalent foundations. arXiv:1211.2851, 2012.
- [Kra15] Nicolai Kraus. *Truncation levels in homotopy type theory*. PhD thesis, University of Nottingham, 2015.

- [LF14] Dan Licata and Eric Finster. Eilenberg–MacLane spaces in homotopy type theory. *LICS*, 2014. <http://dlicata.web.wesleyan.edu/pubs/lf14em/lf14em.pdf>.
- [Lur09] Jacob Lurie. *Higher topos theory*. Number 170 in Annals of Mathematics Studies. Princeton University Press, 2009.
- [Lur14] Jacob Lurie. Higher algebra. Available at <http://www.math.harvard.edu/~lurie/>, September 2014.
- [PW14] Álvaro Pelayo and Michael A. Warren. Homotopy type theory and Voevodsky’s univalent foundations. *Bull. Amer. Math. Soc. (N.S.)*, 51(4):597–648, 2014.
- [Shu14] Michael Shulman. Non-unique splittings of homotopy idempotents. MathOverflow question at <http://mathoverflow.net/questions/189412/non-unique-splittings-of-homotopy-idempotents>, December 2014.
- [Uni13] Univalent Foundations Program. *Homotopy Type Theory: Univalent Foundations of Mathematics*. <http://homotopytypetheory.org/book/>, first edition, 2013.