Sets in homotopy type theory[†]

E G B E R T   R I J K E and B A S   S P I T T E R S

Emails: e.m.rijke@gmail.com, b.a.w.spitters@gmail.com

Received 12 March 2013; revised 25 January 2014

Homotopy type theory may be seen as an internal language for the ∞-category of weak ∞-groupoids. Moreover, weak ∞-groupoids model the univalence axiom. Voevodsky proposes this (language for) weak ∞-groupoids as a new foundation for Mathematics called the univalent foundations. It includes the sets as weak ∞-groupoids with contractible connected components, and thereby it includes (much of) the traditional set theoretical foundations as a special case. We thus wonder whether those 'discrete' groupoids do in fact form a (predicative) topos. More generally, homotopy type theory is conjectured to be the internal language of 'elementary' of ∞-toposes. We prove that sets in homotopy type theory form a ΠW-pretopos. This is similar to the fact that the 0-truncation of an ∞-topos is a topos. We show that both a subobject classifier and a 0-object classifier are available for the type theoretical universe of sets. However, both of these are large and moreover the 0-object classifier for sets is a function between 1-types (i.e. groupoids) rather than between sets. Assuming an impredicative propositional resizing rule we may render the subobject classifier small and then we actually obtain a topos of sets.

1. Introduction

A preliminary version of this paper was ready when the standard reference book on homotopy type theory (The Univalent Foundations Program 2013) was produced. In fact, many of the results of this paper can now also be found in chapter 10 of that book. Conversely, the collaborative writing of that chapter helped us to clarify the presentation of the present article. The paper is also meant to give a readable account of some computer proofs which have meanwhile found their way to https://github.com/HoTT/HoTT/.

Homotopy type theory (Awodey 2012) extends the Curry–Howard correspondence between simply typed λ-calculus, Cartesian closed categories and minimal logic, via extensional dependent type theory, locally Cartesian closed categories and predicate logic (Lambek and Scott 1988; Jacobs 1999) to Martin-Löf type theory with identity types and certain homotopical models. The univalent foundations program (Kapulkin 2012; Pelayo and Warren 2012; The Univalent Foundations Program 2013) extends homotopy type theory with the so-called univalence axiom, thus providing a language for ∞-groupoids. Voevodsky's insight was that this can serve as a new foundation for Mathematics. The ∞-groupoids form the prototypical higher topos (Lurie 2009; Rezk 2010). In fact, homotopy type theory and the univalence axiom can be interpreted in

any such higher topos (Shulman 2013). It is conjectured (Awodey 2012; Shulman 2013) that homotopy type theory with univalence and so-called higher inductive types can provide an 'elementary' definition of a higher topos. In this way, it extends the previous program to use toposes (Johnstone 2002; Mac Lane and Moerdijk 1996) as a foundation of Mathematics.

In the present article, we connect the theory of sets (0-types) in the univalent foundations with the theory of predicative toposes (Moerdijk and Palmgren 2002; van den Berg 2012). The prototypical example of a predicative topos is the category of setoids in Martin-Löf type theory. A setoid (Bishop 1967) is a pair of a type with an equivalence relation on it. In this respect, homotopy type theory may be seen as a generalization of the rich type theory which may be modelled in setoids (Altenkirch 1999; Hofmann 1995). A setoid is also a groupoid in which all hom-sets have at most one inhabitant. Thus groupoids generalize setoids. Hofmann and Streicher showed that groupoids form a model for intensional type theory (Hofmann and Streicher 1998). In their article Hofmann and Streicher propose to investigate whether higher groupoids can also model Martin-Löf type theory; they moreover suggest a form of the univalence axiom for categories: that isomorphic objects be equal. The Streicher/Voevodsky's Kan simplicial set model of type theory (Kapulkin 2012; Streicher 2014) is such a higher dimensional version of Hofmann and Streicher's groupoid model. Moreover, Voevodsky and Streicher recognized that the univalence property holds for Kan simplicial sets. Voevodsky proposed to investigate Martin-Löf type theory with the univalence axiom as a new foundation for Mathematics. Later it was found that the addition of higher inductive types was necessary to model general homotopy colimits.

Grothendieck conjectured that Kan simplicial sets and weak ∞-groupoids are equivalent, however, precisely defining this equivalence is the topic of active research around the 'Grothendieck homotopy hypothesis'. As emphasized by Coquand (Bezem *et al.* 2014), both the 0-truncated weak ∞-groupoids and the 0-truncated Kan simplicial sets are similar to setoids, and constructively so. This is made precise for instance by the fact that the 0-truncation of a model topos is a Grothendieck topos (Rezk 2010, Proposition 9.2) and every Grothendieck topos arises in this way (Rezk 2010, Proposition 9.4). We prove an internal version of the former result. An internal version of the latter result may follow by carrying out the constructive model construction in a (predicative) topos (Bezem *et al.* 2014).

Predicative topos theory follows the methodology of algebraic set theory (Joyal and Moerdijk 1995), a categorical treatment of set theory, which in particular captures the notion of smallness by considering pullbacks of a universally small map. It extends the ideas from the elementary theory of the category of sets (Lawvere and Rosebrugh 2003; Palmgren 2012) by including a universe. It thus seems to be an ideal framework to investigate 0-types. However, there is a catch: under the univalence axiom, the universe of sets is itself not a set but a 1-groupoid. Thus the existing framework of algebraic set theory will not be entirely sufficient, and needs to be revised for univalent purposes. The proper treatment of universes is a main reason for preferring the ∞-groupoid model over the simpler setoid model. Obviously, the possibility to do synthetic homotopy theory is another; see The Univalent Foundations Program (2013).

The paper is organized as follows. In Section 2, we sketch a quick overview of the relevant notions of homotopy type theory on which we rely and we provide the reader with background and precise references to the corresponding results and proofs in The Univalent Foundations Program (2013). We give slight generalizations of definitions and results where that is just as easy as merely giving the results specifically about sets. In doing so, we provide the reader with a sense to what part of the theory is specific about sets and what part is not.

The main body of our article is contained in Section 3. We begin by proving the principle of unique choice if we admit the (-1)-truncation operation, and we show that epimorphisms are surjective. We then prove that Set is a regular category. When we add quotients, Set becomes exact and even a ΠW-pretopos, which is one of the main results of our paper. By adding the univalence axiom, we show that the groupoid Set is a 0-object classifier. In fact, we do this by showing that the universe Type is an object classifier, followed by showing that n-Type is an n-object classifier for every $n : \mathbb{N}$. This is our other main result.

In Section 4, we discuss the representation axiom, the collection axiom and the axiom of multiple choice (AMC). These axioms from algebraic set theory are used in stronger systems for predicative topos theory.

We expect the reader to be familiar with type theory, category theory, algebraic set theory and basic homotopy theory. All background information can be found in The Univalent Foundations Program (2013).

1.1. *Notations, conventions and assumptions*

In this article, we use the notation and conventions developed in The Univalent Foundations Program (2013). We mostly use standard categorical definitions; see Johnstone (2002). There is one difference however: where commuting diagrams in category theory commute with respect to extensional equality, we only require diagrams in type theory to commute up to propositional equality.

As in The Univalent Foundations Program (2013), we use Martin-Löf intensional type theory with a hierarchy of universes á la Russell, universe polymorphism and typical ambiguity. This means that a definition like Set really gives a definition at each universe level. There are only a few places where we need to be explicit about universe levels. The proof that epis are surjective, Theorem 3.10 is one such place where we need both small sets and a large power set. As is common in homotopy type theory, we will freely use the axiom of function extensionality. Using this axiom, all limits can be constructed in type theory (Avigad *et al.* 2012; Rijke and Spitters 2013a). We use higher inductive types to implement pushouts (Lumsdaine and Shulman 2015) and truncations. Although the full details of the computational interpretation of such higher inductive types are being worked out, these concrete instances indeed do have a computational interpretation (Barras 2013; Bezem *et al.* 2014). In our result that Set is a ΠW-pretopos, we will also use the univalence axiom for propositions: for every two (-1)-types P and Q, the function $(P =_{\mathsf{Type}} Q) \to (P \simeq Q)$ is an equivalence. It is also used in Theorem 2.27.

To show that $\mathsf{Type}_\bullet \to \mathsf{Type}$ is an object classifier (Section 3.5) we will use the (full) univalence axiom.

In the previous paragraph, we have used informal categorical terminology such as 'pushout' to refer to what would be interpreted as homotopy pushouts in the ∞-category of weak ∞-groupoids. We will usually omit the word 'homotopy' when we talk about those categorical concepts and we will continue to use this naive ∞-category style without any rigorous claims about their interpretation in the model. We can be more precise when speaking about 1-categories, they will be the pre-categories that are developed in Ahrens *et al.* (2013). Assuming the univalence axiom, Set is in fact a (Rezk complete) 1-category.

2. Preliminaries

2.1. *The very basics of the univalent foundations*

We denote *identity types* $\mathsf{Id}_A(x, y)$ by $x =_A y$ or simply by $x = y$. The *concatenation* of $p : x =_A y$ with $q : y =_A z$ will be denoted by $p \cdot q : x =_A z$. If $P : A \to \mathsf{Type}$ is a family of types over A and $p : x =_A y$ is a path in A and $u : P(x)$, we denote the *transportation of u along p* by $p_*(u) : P(y)$. The type $\sum_{(x:A)} \prod_{(y:A)} (y =_A x)$ witnessing that a type A is *contractible* is written as $\mathsf{isContr}(A)$.

A function $f : A \to B$ is said to be an *equivalence* if there is an element of type

$$\mathsf{isEquiv}(f) :\equiv \left(\sum_{(g:B \to A)} g \circ f \sim \mathsf{id}_A \right) \times \left(\sum_{(h:B \to A)} f \circ h \sim \mathsf{id}_B \right).$$

The *homotopy fibre* $\sum_{(a:A)} f(a) =_B b$ of f at b is denoted by $\mathsf{fib}_f(b)$. A function is an equivalence if and only if all its homotopy fibres are contractible. We write $A \simeq B$ for the type $\sum_{(f:A \to B)} \mathsf{isEquiv}(f)$ and we usually make no notational distinction between an equivalence $e : A \simeq B$ and the underlying function.

By path induction, there is a canonical function assigning an equivalence $A \simeq B$ to every path $p : A =_{\mathsf{Type}} B$. The *univalence axiom* asserts that this function is an equivalence between the types $A =_{\mathsf{Type}} B$ and $A \simeq B$. The principle of function extensionality is a consequence of the univalence axiom.

In the rest of this preliminary chapter, we sketch a quick overview of the theory of homotopy *n*-types and the (co)completeness of those.

2.2. *Introducing the type of sets*

We define the notion of being an *n*-type by recursion on $\mathbb{N}_{-2}$, where $\mathbb{N}_{-2}$ is a version of the natural numbers which starts at -2. We will be mainly concerned with the type 0-Type of all 0-types, which are the sets, but it is not possible to ignore the other values entirely. A concrete reason for this is that 0-Type is itself a 1-type; see Lemma 2.9.

Definition 2.1. A type is said to be a (−2)-*truncated type* if it is contractible. Thus, we define

$$\mathsf{is\text{-}(-2)\text{-}type}(A) :\equiv \mathsf{isContr}(A).$$

For $n : \mathbb{N}_{-2}$, we say that A is $(n+1)$-*truncated* if there is an element of type

$$\text{is-}(n+1)\text{-type}(A) :\equiv \prod_{(x,y:A)} \text{is-}n\text{-type}(x =_A y).$$

We also say that A is an n-*type* if A is n-truncated. We define

$$n\text{-Type} :\equiv \sum_{(A:\text{Type})} \text{is-}n\text{-type}(A).$$

Relying on the interpretation of identity types as path spaces, a useful way of looking at the n-truncated types is that n-types have no interesting homotopical structure above truncation level n.

In the hierarchy of truncatedness we just introduced, we find the sets in 0-Type. Those are the types A with the property that for any two points $x, y : A$, the identity type $x =_A y$ is contractible once inhabited. From a topological point of view, the 0-truncated spaces are those with contractible connected components. Such a space is homotopically equivalent to a discrete space.

Definition 2.2. We introduce the following notations:

$$\text{Prop} :\equiv (-1)\text{-Type} \qquad \text{Set} :\equiv 0\text{-Type}.$$

When $A : \text{Prop}$, we also say that A is a *mere proposition*. Although strictly speaking Prop and Set are dependent pair types, we shall make no notational distinguishment between terms $P : \text{Prop}$ and $A : \text{Set}$ and their underlying types.

The following theorem provides a useful way to show that types are sets. In The Univalent Foundations Program (2013), it is used to prove Hedberg's theorem and to prove that the Cauchy-reals are a set.

Theorem 2.3 (see 7.2.2 in The Univalent Foundations Program (2013)). *Suppose R is a reflexive mere relation on a type X implying identity. Then X is a set, and $R(x, y)$ is equivalent to $x = y$ for all $x, y : X$.*

Proof. Let $\rho : \prod_{(x:X)} R(x, x)$ be a proof of reflexivity of R, and consider a witness $f : \prod_{(x,y:X)} R(x, y) \to (x = y)$ of the assumption that R implies identity. Note first that the two statements in the theorem are equivalent. For on one hand, if X is a set, then $x = y$ is a mere proposition, and since it is logically equivalent to the mere proposition $R(x, y)$ by hypothesis, it must also be equivalent to it. On the other hand, if $x = y$ is equivalent to $R(x, y)$, then like the latter it is a mere proposition for all $x, y : X$, and hence X is a set.

We show that each $f(x, y) : R(x, y) \to x = y$ is an equivalence. By Theorem 4.7.7 in The Univalent Foundations Program (2013), it suffices to show that f induces an equivalence of total spaces:

$$\left(\sum_{(y:X)} R(x, y)\right) \simeq \left(\sum_{(y:X)} x = y\right).$$

The type on the right is contractible, so it suffices to show that the type on the left is contractible too. As the centre of contraction we take the pair $\langle x, \rho(x) \rangle$. It remains to show, for every $y : X$ and every $H : R(x, y)$ that

$$\langle x, \rho(x) \rangle = \langle y, H \rangle,$$

which is by Theorem 2.7.2 of The Univalent Foundations Program (2013) equivalent to

$$\sum_{(p:x=y)} p_*(\rho(x)) = H.$$

But since $R(x, y)$ is a mere proposition, it suffices to show that $x = y$, which we get from $f(H)$. $\square$

2.3. *Closure properties of the n-types*

We list some of the basic properties of n-types; for proofs see The Univalent Foundations Program (2013). Most of the results we cite and mention were proved by Voevodsky Voevodsky (2014) when he introduced the notion of *homotopy levels.* He tried to rationalize the numbering by starting at 0, we follow The Univalent Foundations Program (2013) and the homotopical tradition and start at -2. The proof that n-Type is closed under dependent products requires the function extensionality principle, which is itself a consequence of the univalence axiom. In fact, one does not need to use univalence to show that function extensionality is equivalent to the principle that the (-2)-types are closed under dependent products. The proof that for $n \geqslant -1$, n-Type is also closed under the W type contstructor is a recent result by Danielsson (2013).

For the present paper, the results in Lemmas 2.4–2.9 are of particular interest in the case $n \equiv 0$.

Lemma 2.4 (see Lemma 4.7.3 and Theorems 4.7.4 and 7.1.4 in The Univalent Foundations Program (2013)). *A retract of an n-truncated type is n-truncated. Consequently, n-truncated types are closed under equivalence.*

Lemma 2.5 (see Theorem 7.1.9 in The Univalent Foundations Program (2013)). *Let A be a type and let P be a family of types over A with n-truncated fibres. Then the dependent function type $\prod_{(x:A)} P(x)$ is n-truncated.*

Lemma 2.6 (see Theorem 7.1.8 in The Univalent Foundations Program (2013)). *Let A be n-truncated and that P is a family of n-truncated types over A. Then the dependent pair type $\sum_{(x:A)} P(x)$ is n-truncated.*

Lemma 2.7. *Let A be n-truncated, $n \geqslant -1$ and that $P : A \to$ Type is any family of types over A. Then the well-ordered type $\mathsf{W}_{(x:A)} P(x)$ is n-truncated.*

Lemma 2.8 (see Theorem 7.1.7 in The Univalent Foundations Program (2013)). *If A is n-truncated, then A is $(n + 1)$-truncated. Hence if A is n-truncated, so is $x =_A y$ for each $x, y : A$.*

The second assertion in the following result requires the univalence axiom.

Lemma 2.9 (see Theorem 7.1.11 in The Univalent Foundations Program (2013)). *For any type A and any $n : \mathbb{N}_{-2}$, the type is-n-type(A) is a mere proposition. The type n-Type is itself an $(n + 1)$-truncated type.*

Using higher inductive types, it is possible to implement a left adjoint to the inclusion n-Type $\to$ Type. This left adjoint is called the *n-truncation*; see The Univalent Foundations Program (2013). For the purpose of this paper, we shall only be concerned with the universal property for n-truncation.

Definition 2.10 (a slight extension of Theorem 7.3.2 in The Univalent Foundations Program (2013), see also Theorem 7.7.7)**.** For any type A and any $n : \mathbb{N}_{-2}$ there is an n-truncated type $\|A\|_n$ and a function $|-|_n : A \to \|A\|_n$ such that the function

$$\lambda s.\, s \circ |-|_n : \prod_{(w:\|A\|_n)} P(w) \to \prod_{(a:A)} P(|a|_n)$$

is an equivalence for every $P : \|A\|_n \to n\text{-Type}$.

The (-2)-truncation of a type A is contractible and the (-1)-truncation identifies all elements with each other. Since (-1)-truncation is so common, we will often omit the subscript and write $\|-\|$ instead of $\|-\|_{-1}$. In general, n-truncation maps to n-Type. Therefore, we have for any type A and any $x, y : A$ that the type $|x|_{n+1} =_{\|A\|_{n+1}} |y|_{n+1}$ has all the structure above truncation level n identified. More precisely, we get:

Lemma 2.11 (see Theorem 7.3.12 in The Univalent Foundations Program (2013))**.** *For any type A, any $x, y : A$ and any $n : \mathbb{N}_{-2}$, there is an equivalence*

$$(|x|_{n+1} =_{\|A\|_{n+1}} |y|_{n+1}) \simeq \|x =_A y\|_n.$$

Using the (-1)-truncation $\|-\|$, we can fully implement propositions as (-1)-types.

Definition 2.12. For a type A and (dependent) mere propositions P and Q we define

$$
\begin{aligned}
\top &:\equiv \mathbf{1} & P \Rightarrow Q &:\equiv P \to Q \\
\bot &:\equiv \mathbf{0} & \neg P &:\equiv P \to \mathbf{0} \\
P \wedge Q &:\equiv P \times Q & \forall_{(x:A)} P(x) &:\equiv \prod_{(x:A)} P(x) \\
P \vee Q &:\equiv \|P + Q\| & \exists_{(x:A)} P(x) &:\equiv \left\| \sum_{(x:A)} P(x) \right\|.
\end{aligned}
$$

Recall that bi-implication of (-1)-types implies equivalence of those (-1)-types, which is equivalent by univalence to identity. Notice also that although the product of mere propositions is again a mere proposition, this is not the case for dependent sums. This is the reason why we needed (-1)-truncations to implement propositions as mere propositions.

2.4. *Surjective and injective functions*

The n-truncations are examples of a more general phenomenon called modalities and a large part of the theory of truncations generalizes to arbitrary modalities, as we will show in a forthcoming paper. A very first approximation is available at Rijke and Spitters (2013b).

Definition 2.13. A function $f : A \to B$ is said to be *n-connected* if there is a term of type

$$\prod_{(b:B)} \mathsf{isContr}(\|\mathsf{fib}_f(B)\|_n)$$

Definition 2.14. A function $f : A \to B$ is said to be *n-truncated* if there is a term of type

$$\prod_{(b:B)} \mathsf{is\text{-}}n\text{-type}(\mathsf{fib}_f(b)).$$

The main result about the classes *n*-connected and *n*-truncated functions is that they describe a stable orthogonal factorization system. Every function factors uniquely as an *n*-connected function followed by an *n*-truncated function, see Theorem 7.6.6 in The Univalent Foundations Program (2013) for the precise statement. The unique factorization goes through the *n*-image:

Definition 2.15. Let $f : A \to B$ be a function. We define the *n-image of f* to be the type

$$\mathsf{im}_n(f) :\equiv \textstyle\sum_{(b:B)} \|\mathsf{fib}_f(b)\|_n.$$

In the present article, we are mostly interested in the (-1)-connected and the (-1)-truncated maps, which give factorization of functions through their (-1)-image. We will denote the (-1)-image of a function f simply by $\mathsf{im}(f)$ and call it the *image of f*. Also, it is more customary to talk about surjective and injective functions instead of (-1)-connected and (-1)-truncated functions. We make the following definitions of surjectivity and injectivity, which are equivalent to the definitions of (-1)-connectedness and (-1)-truncatedness respectively.

Definition 2.16. A function $f : A \to B$ is said to be *surjective* if there is a term of type

$$\mathsf{surj}(f) :\equiv \textstyle\prod_{(b:B)} \|\mathsf{fib}_f(b)\|.$$

Definition 2.17. A function $f : A \to B$ is said to be *injective* if there is a term of type

$$\mathsf{inj}(f) :\equiv \textstyle\prod_{(a:A)} \mathsf{isContr}(\mathsf{fib}_f(f(a))).$$

Below, we give the factorization of any function, but we do not go into the details of the uniqueness of such a factorization.

Definition 2.18. Let $f : A \to B$ be a function. Define the functions $\tilde{f} : A \to \mathsf{im}(f)$ and $i_f : \mathsf{im}(f) \to B$ by

$$\tilde{f} :\equiv \lambda a.\, \langle f(a), |(a, \mathsf{refl}_{f(a)})| \rangle$$
$$i_f :\equiv \mathsf{pr}_1.$$

Lemma 2.19. *Let $f : A \to B$ be a function. Then $\tilde{f}$ is surjective and i_f is injective.*

The injective functions are the monomorphisms of Set, which we can also define via a pullback diagram.

Definition 2.20. Let $f : A \to X$ and $g : B \to X$ be functions. We define the homotopy pullback of f and g to be

$$A \times_X B :\equiv \textstyle\sum_{(a:A)} \sum_{(b:B)} (f(a) =_X g(b))$$

and we define $\pi_1 : A \times_X B \to A$ and $\pi_2 : A \times_X B \to B$ to be the projections.

We have the following characterization of $(n+1)$-truncated functions which appears in Rezk (2010) in the setting of model toposes, but not in The Univalent Foundations Program (2013).

Lemma 2.21. *A function $f : A \to B$ is $(n+1)$-truncated if and only if the function*

$$\lambda a.\, \langle a, a, \mathsf{refl}_{f(a)} \rangle : A \to A \times_B A$$

is n-truncated.

Proof. Let $\langle x, y, p \rangle : \sum_{(x,y:A)} f(x) = f(y)$. Then we have a the equivalences

$$\mathsf{fib}_{\lambda a.\, \langle a,a,\mathsf{refl}_{f(a)}\rangle}(\langle x, y, p \rangle) \simeq \sum_{(a:A)} \langle a, a, \mathsf{refl}_{f(a)} \rangle = \langle x, y, p \rangle$$

$$\simeq \sum_{(a:A)} \sum_{(\alpha:a=x)} \sum_{(\beta:a=y)} f(\alpha)^{-1} \cdot f(\beta) = p$$

$$\simeq \sum_{(\beta:x=y)} f(\beta) = p$$

$$\simeq \sum_{(\beta:x=y)} f(\beta)^{-1} \cdot \mathsf{refl}_{f(x)} = p^{-1}$$

$$\simeq \sum_{(\beta:x=y)} \beta_*\big(\mathsf{refl}_{f(x)}\big) = p^{-1}$$

$$\simeq \langle x, \mathsf{refl}_{f(x)} \rangle =_{\mathsf{fib}_f(f(x))} \langle y, p^{-1} \rangle.$$

The latter type is an n-type if and only if $\mathsf{fib}_f(f(x))$ is an $(n+1)$-type. Thus, we see that $\lambda a.\, \langle a, a, \mathsf{refl}_{f(a)} \rangle$ is n-truncated if and only if $\mathsf{fib}_f(f(x))$ is an $(n+1)$-type for each $x : A$. Note that each $\mathsf{fib}_f(f(x))$ is an $(n+1)$-type if and only if each $\mathsf{fib}_f(b)$ is an $(n+1)$-type. □

In particular, a function is injective if and only if the function $A \to A \times_B A$ is an equivalence. In other words, a function is injective precisely when it is a monomorphism, i.e. when the diagram

$$
\begin{array}{ccc}
A & \xrightarrow{\;\mathsf{id}_A\;} & A \\
{\scriptstyle \mathsf{id}_A}\downarrow & & \downarrow{\scriptstyle f} \\
A & \xrightarrow[\;f\;]{} & B
\end{array}
$$

is a pullback diagram.

Let us make the verifications of two of the ingredients of a predicative topos. The first is that sums are disjoint.

Lemma 2.22 (see Theorem 2.12.5 in The Univalent Foundations Program (2013)). *For any two types X and Y we have the equivalences*

$$(\mathsf{inl}(x) =_{X+Y} \mathsf{inl}(x')) \simeq (x =_X x')$$

$$(\mathsf{inl}(x) =_{X+Y} \mathsf{inr}(y)) \simeq \mathbf{0}$$

$$(\mathsf{inr}(y) =_{X+Y} \mathsf{inr}(y')) \simeq (y =_Y y').$$

Consequently, the inclusions $\mathsf{inl} : X \to X + Y$ *and* $\mathsf{inr} : Y \to X + Y$ *are monomorphisms and the diagram*

$$
\begin{array}{ccc}
\mathbf{0} & \longrightarrow & X \\
\downarrow & & \downarrow{\scriptstyle \mathsf{inl}} \\
Y & \xrightarrow[\;\mathsf{inr}\;]{} & X + Y
\end{array}
$$

is a pullback diagram.

We also have the following general result, which has the consequence that Set is *lextensive.*

Theorem 2.23. *Let* $P : A \to \mathsf{Type}$ *be a family of types and let* $f : (\sum_{(a:A)} P(a)) \to B$ *and* $g : X \to B$ *be functions. Then there is an equivalence*

$$\left(\sum_{(a:A)} (P(a) \times_B X)\right) \simeq \left(\sum_{(a:A)} P(a)\right) \times_B X.$$

In particular, there is an equivalence

$$(A_0 \times_B X) + (A_1 \times_B X) \simeq (A_0 + A_1) \times_B X$$

for any three functions $f_0 : A_0 \to B$ *and* $f_1 : A_1 \to B$ *and* $g : X \to B$.

Proof. Note that we have the equivalences

$$
\begin{aligned}
\left(\sum_{(a:A)} (P(a) \times_B X)\right) &\equiv \sum_{(a:A)} \sum_{(u:P(a))} \sum_{(x:X)} (f(a,u) =_B g(x)) \\
&\simeq \sum_{(w:\sum_{(a:A)} P(a))} \sum_{(x:X)} (f(w) =_B g(x)) \\
&\equiv \left(\sum_{(a:A)} P(a)\right) \times_B X. \qquad \square
\end{aligned}
$$

2.5. *Homotopy colimits*

In contrast with limits, general colimits such as quotients are not provided by Martin-Löf type theory. Thus the category of setoids – the left exact completion of the category of types – was considered to work around this deficit. In the univalent foundations, quotients can be introduced as higher inductive types. We present only the results that are essential in the context of sets, a more thorough discussion about higher inductive types can be found either in chapter 6 of The Univalent Foundations Program (2013) or in Rijke and Spitters (2013a).

Definition 2.24. A *(directed) graph* Γ is a pair $\langle \Gamma_0, \Gamma_1 \rangle$ consisting of a type Γ_0 of *points* and a binary relation Γ_1 of *edges*.

Definition 2.25. Let Γ be a graph. We define $\mathsf{colim}(\Gamma)$ to be the higher inductive type with basic constructors

$$
\begin{aligned}
\alpha_0 &: \Gamma_0 \to \mathsf{colim}(\Gamma) \\
\alpha_1 &: \prod_{(i,j:\Gamma_0)} \Gamma_1(i,j) \to \alpha_0(i) = \alpha_0(j).
\end{aligned}
$$

The induction principle for $\mathsf{colim}(\Gamma)$ is that for any family $P : \mathsf{colim}(\Gamma) \to \mathsf{Type}$, if there are

$$
\begin{aligned}
H_0 &: \prod_{(i:\Gamma_0)} P(\alpha_0(i)) \\
H_1 &: \prod_{(i,j:\Gamma_0)} \prod_{(q:\Gamma_1(i,j))} \alpha_1(q)_*(H_0(i)) = H_0(j)
\end{aligned}
$$

then there is a dependent function $f : \prod_{(w:\mathsf{colim}(\Gamma))} P(w)$ with

$$
\begin{aligned}
f(\alpha_0(i)) &:\equiv H_0(i) & \text{for } i : \Gamma_0 \\
\mathsf{apd}_f(\alpha_1(q)) &:= H_1(q) & \text{for } i, j : \Gamma_0 \text{ and } q : \Gamma_1(i,j).
\end{aligned}
$$

These colimits are higher inductive types of the kind that are presented in Section 6.12 in The Univalent Foundations Program (2013), using a binary relation over the type Γ_0 rather than a pair of functions into Γ_0.

Using these higher inductive types, all the homotopy colimits that appear in The Univalent Foundations Program (2013) can be constructed. In particular, we have the pushout of $f : A \to B$ and $g : A \to C$ by considering the graph Γ with

$$\Gamma_0 :\equiv B + C$$
$$\Gamma_1(i, j) :\equiv \mathsf{fib}_\varphi(\langle i, j \rangle)$$

where we take

$$\varphi :\equiv \lambda a. \langle \mathsf{inl}(f(a)), \mathsf{inr}(g(a)) \rangle : A \to (B + C)^2.$$

To see what $\mathsf{fib}_\varphi(-)$ is, note that

$$\mathsf{fib}_\varphi(\langle \mathsf{inl}(b), \mathsf{inr}(c) \rangle) \simeq \sum\nolimits_{(a:A)} (f(a) = b) \times (g(a) = c)$$

and that $\mathsf{fib}_\varphi(-)$ is empty for other combinations of inl and inr. The pushout is equivalently described with the basic constructors

$$\mathsf{inl} : B \to B +_A C$$
$$\mathsf{inr} : C \to B +_A C$$
$$\mathsf{glue} : \prod\nolimits_{(a:A)} \mathsf{inl}(f(a)) = \mathsf{inr}(g(a)).$$

We note that using higher inductive types and univalence for propositions, it is possible to give a new proof of the fact that the axiom of choice, see Equation 3.8.1 in The Univalent Foundations Program (2013), implies the law of excluded middle, see Equation 3.4.1 in The Univalent Foundations Program (2013). Although we will not use this fact in the proof that Set forms a predicative topos, the higher inductive type is an instance of the construction of quotients.

Definition 2.26. Suppose P is a mere proposition. We define the auxiliary binary relation $R_P : \mathbf{2} \to \mathbf{2} \to \mathsf{Type}$ by $R_P(0_{\mathbf{2}}, 1_{\mathbf{2}}) :\equiv P$ and $R_P(b, b') :\equiv \mathbf{0}$ otherwise. Define $\mathbf{2}/P$ to be the type

$$\mathsf{colim}(\langle \mathbf{2}, R_P \rangle).$$

Using Theorem 2.3 it is not hard to see that $\mathbf{2}/P$ is a set, see Lemma 10.1.13 in The Univalent Foundations Program (2013). The basic constructor $\mathbf{2} \to \mathbf{2}/P$ is a surjective function, so we may use the axiom of choice to obtain a section and use the decidability of equality in $\mathbf{2}$ to decide whether P or $\neg P$ holds:

Theorem 2.27 (see Theorem 10.1.14 in The Univalent Foundations Program (2013))**.** *If all surjections between sets merely split, then the law of excluded middle follows.*

Because truncations are left-adjoints, we note that when Γ is a graph, then $\|\mathsf{colim}(\Gamma)\|_0$ is the set-colimit of Γ. To see this, note that the truncation $\|-\|_0$ restricts the universal property of $\mathsf{colim}(\Gamma)$ to only those cases where a family of sets over $\mathsf{colim}(\Gamma)$ is considered.

We end the preliminaries with a discussion on how to take the coequalizer of two functions $f, g : A \to B$ in the univalent category of sets. Its definition as a higher inductive type is straightforward:

Definition 2.28. Let $f, g : A \to B$ be functions between sets. We define the set-coequalizer of f and g to be the type $B/_{f,g} :\equiv \|\mathsf{colim}(\langle B, \mathsf{eq}_{f,g} \rangle)\|_0$, where $\mathsf{eq}_{f,g}$ is the family defined

by

$$\mathsf{eq}_{f,g}(b,b') :\equiv \sum_{(a:A)} (f(a) = b) \times (g(a) = b').$$

We denote the composite function $B \to \mathsf{colim}(\langle B, \mathsf{eq}_{f,g}\rangle) \to B/_{f,g}$ by $c_{f,g}$. A *regular epimorphism* is a function between sets which is (homotopic to) a set-coequalizer.

The following lemma explains that the coequalizer of f and g has indeed the right universal property. This is a general phenomenon; truncated colimits behave as expected.

Lemma 2.29 (see Lemma 10.1.4 in The Univalent Foundations Program (2013)). *Let $f, g : A \to B$ be functions between sets A and B. The set-coequalizer $c_{f,g} : B \to B/_{f,g}$ satisfies the universal property*

$$\prod_{(C:\mathsf{Set})} \prod_{(h:B\to C)} \prod_{(H:h\circ f\sim h\circ g)} \mathsf{isContr}(\sum_{(k:B/_{f,g}\to C)} k \circ c_{f,g} \sim h).$$

3. Set **is a** ΠW-**pretopos**

In this section, we begin by verifying the principle of unique choice. The importance of this result is not in the complication of its proof, but in the absence of the result in some other type theories. In these type theories one introduces a separate sort of 'propositions', which, however, are not necessarily identified with mere propositions. Such an approach may be more general, but less powerful. The principle of unique choice fails in the calculus of constructions (Streicher 1992), in logic enriched type theory (Aczel and Gambino 2002), and in minimal type theory (Maietti and Sambin 2005), the category of prop-valued setoids in CoQ (Spiwack 2011). This principle does hold the model of total setoids using a propositions as types interpretation (Hofmann 1995).

We will show that in the presence of (-1)-truncation, Set becomes a regular category. The natural candidate for coequalizer of the kernel pair of a function is the image of the function. Our proof that the image is indeed the coequalizer is an application of the principle of unique choice. This work is reminiscent of the connections between [ ]-types (Awodey and Bauer 2004), or mono-types (Maietti 2005), and regular categories in an extensional setting.

To show that Set is exact provided that we have quotients, we need to show in addition that every equivalence relation is effective. In other words, given an equivalence relation $R : A \to A \to \mathsf{Prop}$, there is a coequalizer c_R of the pair $\pi_1, \pi_2 : (\sum_{(x,y:A)} R(x,y)) \to A$ and, moreover, the π_1 and π_2 for the kernel pair of c_R.

We consider the pre-category EqRel, which becomes a 1-category by univalence. The pre-category EqRel shares many properties of the pre-category Std of setoids, which is the exact completion of Set. Using higher inductive types and univalence, we will show that we have a quotient functor EqRel $\to$ Set which is moreover left adjoint to the inclusion Set $\to$ EqRel. This adjunction is in general not an equivalence, that would follow from the axiom of choice. With Set being exact, we will be ready to show that Set forms a ΠW-pretopos.

After having shown that Set is exact, we will show that Type has an object classifier. From this we will derive that Set has a subobject classifier. This also shows that if we assume the resizing rule that Prop is equivalent to a type in Set, then Set actually becomes a topos.

3.1. *Regularity of the category of sets*

Definition 3.1. Suppose $P : A \to \mathsf{Type}$ is a family of types over A. We define

$$\mathsf{atMostOne}(P) :\equiv \prod_{(x,y:A)} P(x) \to P(y) \to (x = y)$$

$$\exists!_{(x:A)} P(x) :\equiv (\exists_{(x:A)} P(x)) \times \mathsf{atMostOne}(P).$$

Lemma 3.2. *Suppose that* $P : A \to \mathsf{Prop}$. *If there is an element* $H : \mathsf{atMostOne}(P)$, *then the type* $\sum_{(x:A)} P(x)$ *is a mere proposition.*

Proof. Suppose that $\langle x, u \rangle$ and $\langle x', u' \rangle$ are elements of $\sum_{(x:A)} P(x)$. Then we have the path $p :\equiv H(u, u') : x = x'$. Moreover, there is a path from $p_*(u) = u'$ since $P(x')$ is assumed to be a mere proposition. $\square$

Lemma 3.3. *For any family* $P : A \to \mathsf{Prop}$ *of mere propositions there is a function of type*

$$(\exists!_{(x:A)} P(x)) \to \sum_{(x:A)} P(x).$$

Proof. Suppose we have $H : \mathsf{atMostOne}(P)$ and $K : \exists_{(x:A)} P(x)$. From H it follows that $\sum_{(x:A)} P(x)$ is a mere proposition, and therefore it follows that $\left(\sum_{(x:A)} P(x)\right) \simeq (\exists_{(x:A)} P(x))$. $\square$

Theorem 3.4 (the principle of unique choice). *Suppose that* A *is a type, that* $P : A \to \mathsf{Type}$ *is a family of types over* A *and that* $R : \prod_{(x:A)} (P(x) \to \mathsf{Prop})$ *is a family of mere propositions over* P. *Then there is a function*

$$\left(\prod_{(x:A)} \exists!_{(u:P(x))} R(x, u)\right) \to \sum_{(f:\prod_{(x:A)} P(x))} \prod_{(x:A)} R(x, f(x)).$$

Proof. Suppose that $H : \prod_{(x:A)} (\exists!_{(u:P(x))} R(x))$. By Lemma 3.3 we can find an element of type $\sum_{(u:P(x))} R(x, u)$ for every $x : A$. A function

$$(\prod_{(x:A)} \sum_{(u:P(x))} R(x, u)) \to (\sum_{(f:\prod_{(x:A)} P(x))} \prod_{(x:A)} R(x, f(x)))$$

is obtained from the usual AC_∞. $\square$

The following seemingly stronger variant of $\mathsf{atMostOne}(P)$ helps us showing that $\mathsf{atMostOne}(P)$ is a mere proposition for every $P : A \to \mathsf{Prop}$.

Definition 3.5. Let $P : A \to \mathsf{Type}$ be a family of types over a type A. We define

$$\mathsf{baseLevel}(-1, P) :\equiv \prod_{(x,y:A)} P(x) \to P(y) \to \mathsf{isContr}(x = y).$$

Notice that we could replace $\mathsf{isContr}$ in the definition of $\mathsf{baseLevel}(-1)$ by is-(-2)-type and see that we can easily generalize the notion of $\mathsf{baseLevel}(-1)$ to $\mathsf{baseLevel}(n)$ for $n \geqslant -1$.

Lemma 3.6. *For any* $P : A \to \mathsf{Prop}$, *there is a function of type*

$$\mathsf{atMostOne}(P) \to \mathsf{baseLevel}(-1, P).$$

Proof. We will show that there is a function of type

$$\mathsf{atMostOne}(P) \to \textstyle\prod_{(x,y:A)} P(x) \to P(y) \to \mathsf{isContr}(x=y).$$

Let H be an element of type $\mathsf{atMostOne}(P)$ and let $x, y : A$, $u : P(x)$ and $v : P(y)$. Then we have the terms $\langle x, u \rangle$ and $\langle y, v \rangle$ in $\sum_{(x:A)} P(x)$. Since $\sum_{(x:A)} P(x)$ is a proposition, the path space $\langle x, u \rangle = \langle y, v \rangle$ is contractible. Since P is assumed to be a proposition, there is an equivalence $(x = y) \simeq (\langle x, u \rangle = \langle y, v \rangle)$. Hence it follows that $x = y$ is contractible. $\qquad \square$

Corollary 3.7. *For any family $P : A \to \mathsf{Prop}$ of types, the type $\mathsf{atMostOne}(P)$ is equivalent to $\mathsf{baseLevel}(-1, P)$. In particular, $\mathsf{atMostOne}(P)$ is a mere proposition.*

As an application of unique choice, we show that surjective functions between sets are regular epimorphisms.

Definition 3.8. Let $f : A \to B$ be a function between sets. Define

$$\mathsf{epi}(f) :\equiv \textstyle\prod_{(X:\mathsf{Set})} \prod_{(g,h:B \to X)} (g \circ f \sim h \circ f) \to (g \sim h).$$

Since we have restricted the condition of being an epimorphism to the category of sets, the type $\mathsf{epi}(f)$ is a mere proposition.

Lemma 3.9. *Let f be a function between sets. The following are equivalent:*

  i. *f is an epimorphism.*
 ii. *Consider the pushout diagram.*

$$
\begin{array}{ccc}
A & \xrightarrow{\ f\ } & B \\
\downarrow & & \downarrow{\scriptstyle \iota} \\
1 & \xrightarrow{\ t\ } & \mathcal{C}_f
\end{array}
$$

defining the mapping cone of f. The type $\|\mathcal{C}_f\|_0$ is contractible.
iii. *f is surjective.*

Proof. To show that $\mathsf{epi}(f) \to \mathsf{isContr}(\|\mathcal{C}_f\|_0)$, suppose that $H : \mathsf{epi}(f)$. The basic constructor t of $\mathcal{C}_f$ gives us the element $|t(\star)|_0 : \|\mathcal{C}_f\|_0$. We have to show that

$$\textstyle\prod_{(x:\|\mathcal{C}_f\|_0)} x = |t(\star)|_0.$$

Note that the type $x = |t(\star)|_0$ is a mere proposition because $\|\mathcal{C}_f\|_0$ is a set, hence it is equivalent to show that

$$\textstyle\prod_{(w:\mathcal{C}_f)} |w|_0 = |t(\star)|_0$$

which is by Lemma 2.11 equivalent to

$$\textstyle\prod_{(w:\mathcal{C}_f)} \|w = t(\star)\|.$$

We can use induction on $\mathcal{C}_f$: it suffices to find

$$I_0 : \textstyle\prod_{(b:B)} \|\iota(b) = t(\star)\|$$

$$I_1 : \textstyle\prod_{(a:A)} \mathsf{glue}(a)_*(I_0(f(a))) = |\mathsf{refl}_{t(\star)}|.$$

where $\text{glue} : \prod_{(a:A)} \iota(f(a)) = t(\star)$ is a basic constructor of $\mathcal{C}_f$. Since the type of I_1 is the type of sections of a family of identity types of propositions – which are thus contractible – we get I_1 for free. Since f is epi and since we have $\text{glue} : \iota \circ f \sim (\lambda b.\, t(\star)) \circ f$, we get a homotopy $\iota \sim \lambda b.\, t(\star)$, which gives us I_0.

To show that $\text{isContr}(\|\mathcal{C}_f\|_0) \to \text{surj}(f)$, let $H : \text{isContr}(\|\mathcal{C}_f\|_0)$. Using the univalence axiom, we construct a family $P : \|\mathcal{C}_f\|_0 \to \text{Prop}$ of mere propositions. Note that Prop is a set, so it suffices to define the family $P(\|-\|_0) : \mathcal{C}_f \to \text{Prop}$. For this we can use induction on $\mathcal{C}_f$: we define

$$P(|t(x)|_0) :\equiv \mathbf{1} \qquad\qquad \text{for } x : \mathbf{1}$$
$$P(|\iota(b)|_0) :\equiv \|\text{fib}_f(b)\| \qquad\qquad \text{for } b : B.$$

For $a : A$ the type $\|\text{fib}_f(f(a))\|$ is canonically equivalent to $\mathbf{1}$, which finishes the construction of P. Since $\|\mathcal{C}_f\|_0$ is assumed to be contractible it follows that $P(x)$ is equivalent to $P(|t(\star)|_0)$ for any $x : \|\mathcal{C}_f\|_0$. In particular, we find that $\|\text{fib}_f(b)\|$ is contractible for each $b : B$, showing that f is surjective.

To show that $\text{surj}(f) \to \text{epi}(f)$, let $f : A \to B$ be a surjective function and consider a set C and two functions $g, h : B \to C$ with the property that $g \circ f \sim h \circ f$. Since f is assumed to be surjective, we have an equivalence $B \simeq \text{im}(f)$. Since identity types in sets are propositions, we get

$$\prod_{(b:B)} g(b) = h(b) \simeq \prod_{(w:\text{im}(f))} g(\text{pr}_1 w) = h(\text{pr}_1(w))$$
$$\simeq \prod_{(b:B)} \prod_{(a:A)} \prod_{(p:f(a)=b)} g(b) = h(b)$$
$$\simeq \prod_{(a:A)} g(f(a)) = h(f(a)).$$

By assumption, there is an element of the latter type. $\qquad\qquad\qquad\square$

The proof that epis are surjective in Mines *et al.* (1988) uses the power set operation. This proof can be made predicative by using a large power set and typical ambiguity. A predicative proof for setoids was given by Wilander (2010). The proof above is similar, but avoids setoids by using the pushout and the univalence axiom.

Theorem 3.10. *Surjective functions between sets are regular epimorphisms.*

Proof. Note that it suffices to show that for any function $f : A \to B$, the diagram

$$\sum_{(x,y:A)} f(x) = f(y) \;\underset{\pi_2}{\overset{\pi_1}{\rightrightarrows}}\; A \xrightarrow{\tilde{f}} \text{im}(f)$$

is a coequalizer diagram.

We first construct a homotopy $H : \tilde{f} \circ \pi_1 \sim \tilde{f} \circ \pi_2$. Let $\langle x, y, p \rangle$ be an element of $\sum_{(x,y:A)} f(x) = f(y)$. Then we have $\tilde{f}(\pi_1(\langle x, y, p \rangle)) = \langle f(x), u \rangle$, where u is an element of the contractible type $\|\text{fib}_f(f(x))\|$. Similarly, we have a path $\tilde{f}(\pi_2(\langle x, y, p \rangle)) = \langle f(y), v \rangle$, where v is an element of the contractible type $\|\text{fib}_f(f(y))\|$. Since we have $p : f(x) = f(y)$ and since $\|\text{fib}_f(f(y))\|$ is contractible, it follows that we get a path from $\tilde{f}(\pi_1(\langle x, y, p \rangle))$ to $\tilde{f}(\pi_2(\langle x, y, p \rangle))$, which gives us our homotopy H.

Now suppose that $g : A \to X$ is a function for which there is a homotopy $K : g \circ \pi_1 \sim g \circ \pi_2$. We have to show that the type

$$\sum_{(h : \mathsf{im}(f) \to X)} h \circ \tilde{f} \sim g$$

is contractible. We will apply unique choice to define a function from $\mathsf{im}(f)$ to X. Let $R : \mathsf{im}(f) \to X \to \mathsf{Prop}$ be the relation defined by

$$R(w, x) :\equiv \prod_{(a : A)} (\tilde{f}(a) = w) \to (g(a) = x).$$

There is an element of $\mathsf{atMostOne}(R(w))$ for every $w : \mathsf{im}(f)$. To see this, note that the type $\mathsf{atMostOne}(R(w))$ is a mere proposition. Therefore, there is an equivalence

$$\left(\prod_{(w : \mathsf{im}(f))} \mathsf{atMostOne}(R(w)) \right) \simeq \prod_{(a : A)} \mathsf{atMostOne}(R(\tilde{f}(a))).$$

Let $a : A$, $x, x' : X$, $u : R(\tilde{f}(a), x)$ and $u' : R(\tilde{f}(a), x')$. Then there are the paths $u(a, \mathsf{refl}_{\tilde{f}(a)}) : g(a) = x$ and $u'(a, \mathsf{refl}_{\tilde{f}(a)}) : g(a) = x'$, showing that $x = x'$.

Also, there is an element of $\exists_{(x : X)} R(w, x)$ for every $w : \mathsf{im}(f)$. Indeed, the type

$$\prod_{(w : \mathsf{im}(f))} \exists_{(x : X)} R(w, x)$$

is equivalent to the type

$$\prod_{(a : A)} \exists_{(x : X)} R(\tilde{f}(a), x).$$

The type $\prod_{(a : A)} \sum_{(x : X)} R(\tilde{f}(a), x)$ is inhabited by the element

$$\lambda a. \langle g(a), (\lambda a'. \lambda p. K(\langle a', a, p^{-1} \rangle)) \rangle.$$

This shows that the hypotheses of the principle of unique choice are satisfied, so we get an element of type

$$\sum_{(h : \mathsf{im}(f) \to X)} \prod_{(w : \mathsf{im}(f))} R(w, h(w)).$$

An immediate consequence of the way we constructed our function $h : \mathsf{im}(f) \to X$ is that $h \circ \tilde{f} \sim g$. The result follows now from the observation that the type

$$\sum_{(h' : \mathsf{im}(f) \to X)} h' \circ \tilde{f} \sim h \circ \tilde{f}$$

is contractible because $\tilde{f}$ is an epimorphism. $\qquad\square$

Lemma 3.11. *Pullbacks of surjective functions are surjective. Consequently, pullbacks of coequalizers are coequalizers.*

Proof. Consider a pullback diagram

$$
\begin{array}{ccc}
A & \longrightarrow & B \\
{\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\
C & \underset{h}{\longrightarrow} & D
\end{array}
$$

and assume that g is surjective. Applying the pasting lemma of pullbacks with the morphism $c : \mathbf{1} \to C$, we obtain an equivalence $\mathsf{fib}_f(c) \simeq \mathsf{fib}_g(h(c))$ for any $c : C$. This equivalence gives that f is surjective. $\qquad\square$

Theorem 3.12. *The category* Set *is regular.*

Proof. Set has all limits, so it is finitely complete. Theorem 3.10 gives that the kernel pair of each function has a coequalizer. Lemma 3.11 gives that coequalizers are stable under pullbacks. □

3.2. *The* 1-*category of equivalence relations*

Setoids were introduced by Bishop (1967) to model extensional functions in an unspecified effective framework. Hofmann (1995) developed this theory to build a model of extensional type theory in an intensional type theory. For our purpose the drawback of this generality is that we do not obtain a 1-category of setoids out of such a model of setoids. Hence, we shall restrict to *mere propositional* relations to obtain the pre-category of setoids and we will end up with a category of equivalence relations and respectful functions on their carriers as morphisms.

Due to the restriction we have to put on the 'setoids' to form a 1-category, Set becomes a reflective subcategory of the category EqRel of setoids with respectful functions; see Section 3.3. In the presence of the axiom of choice (The Univalent Foundations Program 2013, Equation 3.8.1) the categories are even equivalent.

The construction in Section 3.3 is reminiscent of the ex/lex completion. It uses setoids with respectful functions. Instead of functions one could also consider total functional relations. Since Set is a regular 1-category, these choices are equivalent. The construction in Section 3.4 uses a Yoneda construction also reminiscent of the ex/lex completion; see Maietti and Rosolini (2012) for an overview.

Objects of EqRel consist of types accompanied with an equivalence relation. Since we want to end up with a 1-category, we shall have to require that the underlying type of a setoid is always 0-truncated, i.e. a set. In the proof relevant definition of setoids, an equivalence relation over a type A is a family $R : A \to A \to$ Type which is reflexive, symmetric and transitive. Also here we have to compromise: we will take proposition-valued equivalence relations instead.

Definition 3.13. An *equivalence relation* over a type A consists of a mere relation $R : A \to A \to$ Prop which is *reflexive*, *symmetric* and *transitive*, i.e. there are elements

$$\rho : \prod_{(x:A)} R(x,x)$$
$$\sigma : \prod_{(x,y:A)} R(x,y) \to R(y,x)$$
$$\tau : \prod_{(x,y,z:A)} R(y,z) \to R(x,y) \to R(x,z).$$

We also write isEqRel(R) for the type witnessing that R is an equivalence relation.

Definition 3.14. We define

$$\mathrm{ob}(\mathrm{EqRel}) :\equiv \sum_{(A:\mathrm{Set})} \sum_{(R:A\to A\to\mathrm{Prop})} \mathrm{isEqRel}(R).$$

Usually we shall slightly abuse notation and speak of $\langle A, R \rangle$ as an object of EqRel, leaving the witness of isEqRel(R) implicit.

Definition 3.15. Let $\Gamma :\equiv \langle A, R \rangle$ and $\Delta :\equiv \langle B, S \rangle$ be objects of EqRel. A morphism f from Δ to Γ is a pair $\langle f_0, f_1 \rangle$ consisting of

$$f_0 : B \to A$$
$$f_1 : \prod_{(x,y:B)} S(x,y) \to R(f_0(x), f_0(y)).$$

Thus, we define

$$\mathsf{hom}(\Delta, \Gamma) :\equiv \sum_{(f_0:B\to A)} \prod_{(x,y:B)} S(x,y) \to R(f_0(x), f_0(y)).$$

Which is a sub*set* of the function *set*. We will usually denote the type $\mathsf{hom}(\Delta, \Gamma)$ by $\Delta \to \Gamma$. The identity morphisms id_Γ and the composite morphisms $g \circ f$ are defined in the obvious way.

In the following theorem, we use the univalence axiom to deduce that we get a 1-category of equivalence relations. However, in our result that Set is a ΠW-pretopos we will not use univalence and hence we will not use the fact that EqRel is a 1-category.

Lemma 3.16. *For any two setoids Γ and Δ, the type of isomorphisms from Δ to Γ is equivalent to the type $\Delta = \Gamma$ in* ob(EqRel).

Proof. First observe that we have an equivalence

$$(\Delta = \Gamma) \simeq \left(\sum_{(e:B\simeq A)} \prod_{(x,y:A)} S(e^{-1}(x), e^{-1}(y)) \simeq R(x,y) \right).$$

As for the type of isomorphisms from Δ to Γ, note that we have an equivalence

$$(\Delta \cong \Gamma) \simeq \Big(\sum_{(\langle f_0, g_0, \eta_0, \epsilon_0 \rangle : B \cong A)}$$
$$\left(\prod_{(x,y:B)} S(x,y) \to R(f_0(x), f_0(y)) \right)$$
$$\times \left(\prod_{(x,y:A)} R(x,y) \to S(g_0(x), g_0(y)) \right) \Big).$$

Since B and A are assumed to be sets, we have that $(B \cong A) \simeq (B \simeq A)$ and therefore it suffices to show that the type

$$\left(\prod_{(x,y:B)} S(x,y) \to R(f_0(x), f_0(y)) \right) \times \left(\prod_{(x,y:A)} R(x,y) \to S(g_0(x), g_0(y)) \right)$$

is equivalent to the type $\prod_{(x,y:A)} (S(g_0(x), g_0(y)) \simeq R(x,y))$ for every isomorphism $\langle f_0, g_0, \eta_0, \epsilon_0 \rangle : B \cong A$. Note that both types are mere propositions, so we only have to find implications in both directions. These can be found by using $\eta_0 : g_0 \circ f_0 \sim \mathsf{id}_B$ and $\epsilon_0 : f_0 \circ g_0 \sim \mathsf{id}_A$. □

We already mentioned the inclusion of Set into EqRel. We define it on objects by $A \mapsto \langle A, \mathsf{Id}_A \rangle$. Recall that for each function $f : A \to B$ there is a function $\mathsf{ap}_f : \prod_{(x,y:A)} (x = y) \to (f(x) = f(y))$, and hence defines a map between the setoids $\langle A, \mathsf{Id}_A \rangle$ and $\langle B, \mathsf{Id}_B \rangle$. It comes as no surprise that this determines a functor R.

Lemma 3.17. *The inclusion of* Set *into* EqRel *is full and faithful.*

Proof. We have to show that for any two sets A and B, the inclusion determines an equivalence $(A \to B) \simeq (\langle A, \mathsf{Id}_A \rangle \to \langle B, \mathsf{Id}_B \rangle)$. Naturally, we choose the inverse to be the first projection. Since the identity types on sets are mere propositions, it is immediate that the first projection is a section for the inclusion. □

3.3. *Quotients*

Given an object $\langle A, R \rangle$ of EqRel, we wish to define a set $Q(A, R)$ together with a morphism $f : \langle A, R \rangle \to \langle Q(A, R), \mathsf{Id}_{Q(A,R)} \rangle$ in EqRel with the universal property that precomposition with f_0 gives an equivalence

$$(Q(A, R) \to Y) \simeq \mathsf{hom}_{\mathsf{EqRel}}(\langle A, R \rangle, \langle Y, \mathsf{Id}_Y \rangle)$$

for every set Y. In other words, we are looking for a left adjoint to the inclusion $X \mapsto \langle X, \mathsf{Id}_X \rangle$ of Set into EqRel. Such a left adjoint is mapping the setoid $\langle A, R \rangle$ to the quotient A/R.

There are several solutions to this problem, of which we present two. The first solution uses higher inductive types of the kind presented in Definition 2.25.

Definition 3.18. Let A be a set and let $R : A \to A \to \mathsf{Prop}$ be a binary mere relation over A (not necessarily an equivalence relation). We define A/R to be the type $\|\mathsf{colim}(\langle A, R \rangle)\|_0$.

Since a binary relation $R : A \to A \to \mathsf{Type}$ is equivalently described as a pair of functions by the two projections $\pi_1, \pi_2 : \left(\sum_{(x,y:A)} R(x, y) \right) \to A$, we get the following lemma from Lemma 2.29:

Lemma 3.19. *Let A be a set and let $R : A \to A \to \mathsf{Prop}$ be a binary mere relation over A. Then A/R is the (set-)coequalizer of the two projections $\pi_1, \pi_2 : \left(\sum_{(x,y:A)} R(x, y) \right) \to A$.*

Using the induction principle of each A/R, we can extend the function $\lambda \langle A, R \rangle. A/R$ to a functor Q from EqRel to Set in a canonical way.

We check that quotients have the expected universal properties.

Theorem 3.20. *The functor Q is left adjoint to the inclusion $i : \mathsf{Set} \to \mathsf{EqRel}$. Thus, Set is a reflective subcategory of EqRel.*

Proof. We have to show that there are

 i. a unit $\eta : \mathbf{1} \to i \circ Q$.
 ii. and a counit $\varepsilon : Q \circ i \to \mathbf{1}$
iii. satisfying the triangle identities

$$\varepsilon_{A/R} \circ Q(\eta_{(A,R)}) = \mathsf{refl}_{A/R} \qquad \text{and} \qquad i(\varepsilon_A) \circ \eta_{\langle A, \mathsf{Id}_A \rangle} = \mathsf{refl}_{\langle A, \mathsf{Id}_A \rangle}.$$

For the unit we take $\eta_{(A,R),0} :\equiv c_R$ and $\eta_{(A,R),1} :\equiv p_R$, where c_R is the coequalizer of the pair $\pi_1, \pi_2 : (\sum_{(x,y:A)} R(x, y)) \to A$ and where

$$p_R : \prod_{(x,y:A)} R(x, y) \to (c_R(x) = c_R(y))$$

is also a basic constructor of A/R.

For the counit note that the canonical constructor $c_{\mathsf{Id}_A} : A \mapsto A/\mathsf{Id}_A$ of A/Id_A is an equivalence. Hence we define $\varepsilon_A : A/\mathsf{Id}_A \to A$ to be $c_{\mathsf{Id}_A}^{-1}$.

Note that $Q(\eta_{(A,R)}) : A/R \to (A/R)/\mathsf{Id}_{A/R}$ is the unique map with the the property that the square

$$
\begin{array}{ccc}
A & \xrightarrow{\;c_R\;} & A/R \\
{\scriptstyle c_R}\Big\downarrow & & \Big\downarrow{\scriptstyle c_{\mathsf{Id}_{A/R}}} \\
A/R & \xrightarrow[Q(\eta(A,R))]{} & (A/R)/\mathsf{Id}_{A/R}
\end{array}
$$

and therefore we have a homotopy $Q(\eta_{(A,R)}) \sim c_{\mathsf{Id}_{A/R}}$. By definition we have $\varepsilon_{A/R} :\equiv c_{\mathsf{Id}_{A/R}}^{-1}$, hence the triangle identity $\varepsilon_{A/R} \circ Q(\eta_{(A,R)}) = \mathsf{refl}_{A/R}$ follows.

For the other triangle equality, note that the pair $i(\varepsilon_A)$ consists of the function $i(\varepsilon_A)_0 :\equiv c_{\mathsf{Id}_A}^{-1}$ and the function $i(\varepsilon_A)_1$ which is the canonical proof that equivalences preserve path relations. We also have $\eta_{(A,\mathsf{Id}_A)}$ given by the function $\eta_{(A,\mathsf{Id}_A),0} :\equiv c_{\mathsf{Id}_A}$ and the basic constructor $\eta_{(A,\mathsf{Id}_A),1} :\equiv p_{\mathsf{Id}_A}$ witnessing that $\eta_{(A,R),0}$ preserves the path relation. Since A/Id_A is a set, it follows that $\eta_{(A,\mathsf{Id}_A)}$ is just the canonical proof that $\eta_{(A,\mathsf{Id}_A)}$ preserves the path relation and hence we get the other triangle equality. $\qquad\square$

To prove that this is an equivalence of categories we would need to show that $\langle A/R, \mathsf{Id}_{A/R}\rangle$ and $\langle A, R\rangle$ are isomorphic setoids. It requires AC_{-1} to obtain a section for the surjective map $A \to A/R$.

Definition 3.21. A mere relation $R : A \to A \to \mathsf{Prop}$ is said to be *effective* if the square

$$
\begin{array}{ccc}
\sum_{(x,y:A)} R(x,y) & \xrightarrow{\;\pi_1\;} & A \\
{\scriptstyle \pi_2}\Big\downarrow & & \Big\downarrow{\scriptstyle c_R} \\
A & \xrightarrow[\;c_R\;]{} & A/R
\end{array}
$$

is a pullback square.

The following proposition uses univalence for mere propositions.

Theorem 3.22. *Suppose* $\langle A, R\rangle$ *is an object of* EqRel. *Then there is an equivalence*

$$(c_R(x) = c_R(y)) \simeq R(x,y)$$

for any $x, y : A$. *In other words, equivalence relations are effective.*

Proof. We begin by extending R to a mere relation $\tilde{R} : A/R \to A/R \to \mathsf{Prop}$. After the construction of $\tilde{R}$ we will show that $\tilde{R}$ is equivalent to the identity type on A/R. We define $\tilde{R}$ by double induction on A/R (note that Prop is a set by univalence for mere propositions). We define $\tilde{R}(c_R(x), c_R(y)) :\equiv R(x,y)$. For $r : R(x,x')$ and $s : R(y,y')$, the transitivity and symmetry of R gives an equivalence from $R(x,y)$ to $R(x',y')$. This completes the definition of $\tilde{R}$. To finish the proof of the proposition, we need to show that $\tilde{R}(w,w') \simeq (w = w')$ for every $w, w' : A/R$. We can do this by showing that the type $\sum_{(w':A/R)} \tilde{R}(w,w')$ is contractible for each $w : A/R$. We do this by induction. Let $x : A$. We have the element $\langle c_R(x), \rho(x)\rangle : \sum_{(w':A/R)} \tilde{R}(c_R(x), w')$, where ρ is the reflexivity term of R, hence we only have to show that

$$\prod_{(w':A/R)} \prod_{(r:\tilde{R}(c_R(x),w'))} \langle w', r\rangle = \langle c_R(x), \rho(x)\rangle,$$

which we do by induction on w'. Let $y : A$ and let $r : R(x, y)$. Then we have the path $p_R(r)^{-1} : c_R(y) = c_R(x)$. We automatically get a path from $p_R(r)^{-1}{}_*(r) = \rho(x)$, finishing the proof. $\qquad\square$

3.4. *Voevodsky's impredicative quotients*

A second construction of quotients is due to Voevodsky (Pelayo *et al.* 2013). He defined the quotient A/R as the type of equivalence classes of R, i.e. as the image of R in $A \to \mathsf{Prop}$. This gives a direct construction of quotients, but it requires a resizing rule. In this section we treat Voevodsky's construction of quotients, but we note up front that our result that Set is a ΠW-pretopos does not rely on the material presented here. Throughout this section, we assume that $\langle A, R \rangle$ is an object of EqRel.

Definition 3.23. A predicate $P : A \to \mathsf{Prop}$ is said to be an *equivalence class* with respect to R if there is an element of type

$$\mathsf{isEqClass}(R, P) :\equiv \sum_{(x:A)} \prod_{(y:A)} R(x, y) \simeq P(y).$$

Definition 3.24. We define

$$A /\!\!/ R :\equiv \sum_{(P:A \to \mathsf{Prop})} \|\mathsf{isEqClass}(R, P)\|.$$

Using univalence for mere propositions, the following is a consequence of the definition:

Lemma 3.25. *The type $A /\!\!/ R$ is equivalent to $\mathsf{im}(R)$, the image of $R : A \to (A \to \mathsf{Prop})$.*

In Theorem 3.10, we have shown that images are coequalizers. In particular, we immediately get the coequalizer diagram

$$\sum_{(x,y:A)} R(x) = R(y) \overset{\pi_1}{\underset{\pi_2}{\rightrightarrows}} A \longrightarrow A /\!\!/ R$$

We can use this to show that any equivalence relation is effective.

Theorem 3.26. *Let $f : A \to B$ between any two sets. Then the relation $\ker(f) : A \to A \to \mathsf{Type}$ given by $\ker(f, x, y) :\equiv f(x) = f(y)$ is effective.*

Proof. We will use that $\mathsf{im}(f)$ is the coequalizer of $\pi_1, \pi_2 : (\sum_{(x,y:A)} f(x) = f(y)) \to A$; we get this equivalence from Theorem 3.10. Note that the canonical kernel pair of the function $\mathcal{C}_f :\equiv \lambda a. \langle f(a), \tau_1(\langle a, \mathsf{refl}_{f(a)} \rangle) \rangle$ consists of the two projections

$$\pi_1, \pi_2 : (\sum_{(x,y:A)} \mathcal{C}_f(x) = \mathcal{C}_f(y)) \to A.$$

For any $x, y : A$, we have equivalences

$$\mathcal{C}_f(x) = \mathcal{C}_f(y) \simeq \sum_{(p:f(x)=f(y))} p_*(\tau_1(\langle x, \mathsf{refl}_{f(x)} \rangle)) = \simeq \tau_1(\langle y, \mathsf{refl}_{f(x)} \rangle)$$
$$\simeq f(x) = f(y),$$

where the last equivalence holds because $\|\mathsf{fib}_f(b)\|$ is a mere proposition for any $b : B$. Therefore, we get that

$$(\sum_{(x,y:A)} \mathcal{C}_f(x) = \mathcal{C}_f(y)) \simeq (\sum_{(x,y:A)} f(x) = f(y))$$

and hence we may conclude that ker f is an effective relation. $\square$

Theorem 3.26 show again that equivalence relations are effective.

Theorem 3.27. *Equivalence relations are effective and $A/R \simeq A /\!\!/ R$.*

Proof. We need to analyse the coequalizer diagram

$$\sum\nolimits_{(x,y:A)} R(x) = R(y) \underset{\pi_2}{\overset{\pi_1}{\rightrightarrows}} A \longrightarrow A /\!\!/ R$$

By the univalence axiom, the type $R(x) = R(y)$ is equivalent to the type of homotopies from $R(x)$ to $R(y)$, which is equivalent to $\prod_{(z:A)} R(x, z) \simeq R(y, z)$. Since R is an equivalence relation, the latter type is equivalent to $R(x, y)$. To summarize, we get that $(R(x) = R(y)) \simeq R(x, y)$, so R is effective since it is equivalent to an effective relation. Also, the diagram

$$\sum\nolimits_{(x,y:A)} R(x, y) \underset{\pi_2}{\overset{\pi_1}{\rightrightarrows}} A \longrightarrow A /\!\!/ R$$

is a coequalizer diagram. Since coequalizers are unique up to equivalence, it follows that $A/R \simeq A /\!\!/ R$. $\square$

One may wonder about the predicative interpretation of the quotient constructions above. One could argue that the construction using higher inductive types is predicative by considering the interpretation of this quotient in the setoid model (Altenkirch 1999; Barras 2013). In this model, the quotient does not raise the universe level. A similar observation holds for constructions that can be carried out in the groupoid model (Hofmann and Streicher 1998). These observations should suffice for the set-level higher inductive types we use in the present paper.

We have an inclusion $\mathsf{Prop}_{\mathsf{Type}_i} \to \mathsf{Prop}_{\mathsf{Type}_{i+1}}$. The assumption that this map is an equivalence is called the *propositional resizing axiom*; see The Univalent Foundations Program (2013). This form of impredicativity would make Voevodsky's quotient small.

The following replacement axiom is derivable from the propositional resizing axiom; see Voevodsky (2012).

Lemma 3.28. *Let Type be a universe and $X : \mathsf{Type}$, if $f : X \twoheadrightarrow Y$ is a surjection, Y is a set then there exists a $Z : \mathsf{Type}$ which is equivalent to Y.*

Proof. Define $Z :\equiv X /\!\!/ \ker f$ using a map to the small mere propositions in Definition 3.24. Then $Z : \mathsf{Type}$ and $Z :\equiv X / \ker f \simeq \mathsf{im} f \simeq Y$. $\square$

3.5. *The object classifier*

One of the reasons that the definition of predicative topos in van den Berg (2012) contains such a long list of requirements is the absence of a small subobject classifier. Nevertheless, in Martin-Löf type theory we have the possibility of considering a tall hierarchy of universes nested in one another according to the ordering of the hierarchy. While a universe with a subobject classifier would be impredicative, there is no problem with subobject classifiers at higher universe levels, or simply large subobject classifiers.

In higher topos theory one considers not only subobject classifiers, which classify the monomorphisms, but also object classifiers which classify more general classes of maps. In this section, we will establish the existence of an internal analogue of such large object classifiers. They will always sit a higher universe. Moreover, we will find an n-object classifier for every $n : \mathbb{Z}_{\geqslant -2}$, where the n-object classifier will classify the functions with n-truncated homotopy fibres. In Section 2, we saw that the monomorphisms are exactly the (-1)-truncated functions. Therefore, the (-1)-object classifier will correspond to a (large) subobject classifier.

In addition to the size issue of the object classifiers, we will see that the n-object classifier will generally not be n-Type, but an $n + 1$-Type. This observation should be regarded in contrast to the theory of predicative toposes, where a universal small map is required to exist. Such a universal small map is suggested to be a map between sets, but it seems that within the current setting of homotopy type theory we cannot expect such a map to exist. The main reason is that the universal small map of sets will in general be a map of groupoids; a universal small map of groupoids will in general be a map of 2-groupoids, etc.

Theorem 3.29. *For any type B there is an equivalence*

$$\chi : \left(\textstyle\sum_{(A:\mathsf{Type})} A \to B\right) \simeq B \to \mathsf{Type}.$$

Likewise, there is an equivalence

$$\chi_n : \left(\textstyle\sum_{(A:\mathsf{Type})} \sum_{(f:A\to B)} \prod_{(b:B)} \mathsf{is}\text{-}n\text{-}\mathsf{type}(\mathsf{fib}_f(b))\right) \simeq (B \to n\text{-}\mathsf{Type})$$

for every $n : \mathbb{Z}_{\geqslant -2}$.

Proof. We begin by constructing the first equivalence, i.e. we have to construct functions

$$\chi : \left(\textstyle\sum_{(A:\mathsf{Type})} A \to B\right) \to B \to \mathsf{Type}$$
$$\psi : \left(B \to \mathsf{Type}\right) \to \left(\textstyle\sum_{(A:\mathsf{Type})} A \to B\right).$$

The function χ is defined by $\chi(\langle A, f\rangle, b) :\equiv \mathsf{fib}_f(b)$. The function ψ is defined by $\psi(P) :\equiv \langle (\sum_{(b:B)} P(b)), \mathsf{pr}_1 \rangle$. Now we have to verify that $\chi \circ \psi \sim \mathsf{id}$ and that $\psi \circ \chi \sim \mathsf{id}$:

  i. Let P be a family of types over B. It is a basic fact (The Univalent Foundations Program 2013) that $\mathsf{fib}_{\mathsf{pr}_1}(b) \simeq P(b)$ and therefore it follows immediately that $P \sim \chi(\psi(P))$.
 ii. Let $f : A \to B$ be a function. We have to find a path

$$\langle (\textstyle\sum_{(b:B)} \mathsf{fib}_f(b)), \mathsf{pr}_1 \rangle = \langle A, f\rangle$$

First note that we have the basic equivalence $e : \left(\sum_{(b:B)} \mathsf{fib}_f(b)\right) \simeq A$ with $e(b, a, p) :\equiv a$ and $e^{-1}(a) :\equiv \langle f(a), a, \mathsf{refl}_{f(a)}\rangle$. It also follows that $e_*(\mathsf{pr}_1) = \mathsf{pr}_1 \circ e^{-1}$. From this, we immediately read off that $(e_*(\mathsf{pr}_1))(a) = f(a)$ for each $a : A$. This completes the proof of the first of the asserted equivalences.

To find the second set of equivalences, note that if we restrict χ to functions with n-truncated homotopy fibres we get a family of n-truncated types. Likewise, if we restrict ψ to a family of n-truncated types we get a function with n-truncated homotopy fibres. To finish the proof we observe that truncatedness is a mere proposition, hence adding it

as a restriction on both sides does not disturb the fact that the two functions are inverse equivalences. □

Definition 3.30. Define

$$\mathsf{Type}_\bullet :\equiv \sum_{(A:\mathsf{Type})} A \qquad \text{and} \qquad (n\text{-}\mathsf{Type})_\bullet :\equiv \sum_{(A:n\text{-}\mathsf{Type})} A.$$

Thus, $\mathsf{Type}_\bullet$ stands for the *pointed types* (by analogy with the pointed spaces) and $(n\text{-}\mathsf{Type})_\bullet$ stands for the pointed n-types.

The following theorem states that we have an object classifier.

Theorem 3.31. *Let $f : A \to B$ be a function. Then the diagram*

$$
\begin{array}{ccc}
A & \xrightarrow{\vartheta_f} & \mathsf{Type}_\bullet \\
{\scriptstyle f}\downarrow & & \downarrow{\scriptstyle \mathsf{pr}_1} \\
B & \xrightarrow{\chi_f} & \mathsf{Type}
\end{array}
$$

is a pullback diagram. Here, the function ϑ_f is defined by

$$\lambda a. \langle \mathsf{fib}_f(f(a)), \langle a, \mathsf{refl}_{f(a)} \rangle \rangle.$$

A similar statement holds when we replace Type by n-Type.

Proof. Note that we have the equivalences

$$
\begin{aligned}
A &\simeq \sum_{(b:B)} \mathsf{fib}_f(b) \\
&\simeq \sum_{(b:B)} \sum_{(X:\mathsf{Type})} \sum_{(p:\mathsf{fib}_f(b)=X)} X \\
&\simeq \sum_{(b:B)} \sum_{(X:\mathsf{Type})} \sum_{(x:X)} \mathsf{fib}_f(b) = X \\
&\equiv B \times_{\mathsf{Type}} \mathsf{Type}_\bullet.
\end{aligned}
$$

which gives us a composite equivalence $e : A \simeq B \times_{\mathsf{Type}} \mathsf{Type}_\bullet$. We may display the action of this composite equivalence step by step by

$$
\begin{aligned}
a &\mapsto \langle f(a), \langle a, \mathsf{refl}_{f(a)} \rangle \rangle \\
&\mapsto \langle f(a), \mathsf{fib}_f(f(a)), \mathsf{refl}_{\mathsf{fib}_f(f(a))}, \langle a, \mathsf{refl}_{f(a)} \rangle \rangle \\
&\mapsto \langle f(a), \mathsf{fib}_f(f(a)), \langle a, \mathsf{refl}_{f(a)} \rangle, \mathsf{refl}_{\mathsf{fib}_f(f(a))} \rangle
\end{aligned}
$$

Therefore, we get homotopies $f \sim \pi_1 \circ e$ and $\vartheta_f \sim \pi_2 \circ e$. □

Lemma 3.32. *The type $\mathsf{Prop}_\bullet$ is contractible.*

Proof. Suppose that $\langle P, u \rangle$ is an element of $\sum_{(P:\mathsf{Prop})} P$. Then we have $u : P$ and hence there is an element of type $\mathsf{isContr}(P)$. It follows that $P \simeq \mathbf{1}$ and therefore we get from the univalence axiom that there is a path $\langle P, u \rangle = \langle \mathbf{1}, \star \rangle$. □

If we use the resizing rules we can replace the large type $\sum_{(P:\mathsf{Prop})} P$ with the small type $\mathbf{1}$. Moreover, resizing makes Prop small, so in this way we would obtain the usual notion of a small subobject classifier. Without the resizing rules we do obtain a large subobject classifier.

3.6. Set *is a* ΠW-*pretopos*

We assume the existence of the higher inductive types for truncation and quotients. The univalence axiom is used in Lemma 3.9, but not to prove that surjections are epimorphisms. We do use propositional univalence in Theorem 3.10.

Theorem 3.33. *The category* Set *is a* ΠW-*pretopos.*

Proof. We have an initial object, disjoint finite sums Lemma 2.22, and finite limits (Definition 2.20). Sums are stable under pullback; see Theorem 2.23. So, Set is lextensive. Set is locally Cartesian closed. This follows from the preparations we made in Section 2, using the fact that the existence of Π-types (and functional extensionality) gives local Cartesian closure e.g. Jacobs (1999, Proposition 1.9.8). The category Set is regular (Theorem 3.12) and quotients are effective (Theorem 3.22). We thus have an exact category, since it is also lextensive, we have a pretopos. It has Π-types (Lemma 2.5) and W-types (Lemma 2.7), so we have a ΠW-pretopos. □

One wonders what prevents Set from being a topos. We lack the impredicativity to define the subobject classifier. If we assume the resizing rules from Section 3.4, then Prop becomes small. We have seen in Theorem 3.31 that it satisfies the properties of a subobject classifier and hence we actually obtain a topos: Set is also (locally) Cartesian closed and has all finite limits and colimits.

4. Choice and collection axioms

In this section we study two axioms, the axiom of collection and the AMC, that extend beyond the basic set theory. It seems that these axioms are not provable in homotopy type theory, but we investigate their relationship with the other axioms. As a first attempt, we would try to investigate these axioms in the framework for algebraic set theory. However, in homotopy type theory the universe of sets is a groupoid and hence none of the 1-categorical frameworks quite satisfy our needs. Thus we follow algebraic set theory only loosely. We will work in the naive '∞-category' of types, but observe that most of the constructions either deal with sets or with maps that have set-fibres. In these cases, the constructions align with the constructions in the 1-category Set. In particular, we use (-1)-connected map for cover and use (homotopy) pullbacks. We note that such a pullback between sets reduces to the 1-pullback in the 1-category Set.

In algebraic set theory, one considers a category of classes and isolates the sets within them. In the present paper, next to this size issue, we are mostly concerned with the dimension, the complexity of the equality. Although the results in this section may not be spectacular, we record to what extent the seemingly natural framework of algebraic set theory works and where it breaks down.

4.1. *Class of stable maps*

Definition 4.1. A class $\mathcal{S} : \prod_{(X,Y:\mathsf{Type})} (X \to Y) \to \mathsf{Prop}$ is *stable* (Moerdijk and Palmgren 2002, Definition 3.1) if it satisfies:

Pullback stability. In a pullback diagram as below, we have $\mathcal{S}(f) \to \mathcal{S}(g)$.

$$
\begin{array}{ccc}
X & \longrightarrow & A \\
{\scriptstyle g}\downarrow & & \downarrow{\scriptstyle f} \\
Y & \underset{h}{\longrightarrow} & B
\end{array}
$$

Descent. If h in the above diagram is surjective, then $\mathcal{S}(g) \to \mathcal{S}(f)$.

Sum. If $\mathcal{S}(f)$ and $\mathcal{S}(g)$, then $\mathcal{S}(f + g)$. A class of stable maps is *locally full* (Moerdijk and Palmgren 2002, 3.2) if for all $g : X \to Y$ and $f : Y \to Z$ such that $\mathcal{S}(f) : \mathcal{S}(g)$ iff $\mathcal{S}(fg)$.

A class of maps $\mathcal{S}$ is called a *class of small maps* if it is stable, locally full and for each $X, \mathcal{S}_X$ – the small maps over X – forms a ΠW-pretopos see Moerdijk and Palmgren (2002, 3.3).

Theorem 4.2. *The class of set-fibred maps is a class of small maps.*

Proof. The class of maps with set-fibres is stable. It even has dependent sums.

We claim that it is locally full: if g has set-fibres, then $f \circ g$ has set-fibres, as sets are closed under Σ-types. Conversely, fix $y \in Y$, then $\mathrm{fib}_g(y)$ is the pullback

$$
\sum\nolimits_{(x : \mathrm{fib}_{f \circ g}(f(y)))} (g(x) = y).
$$

By the use of the object classifier, Theorem 3.31, we see that the type $\mathcal{S}_X$ is equivalent to the sets in context X. Now, sets in any context form a ΠW-pretopos; see Section 3. $\square$

4.2. *Representable classes of small maps*

Definition 4.3. A commuting diagram of the form

$$
\begin{array}{ccc}
X & \longrightarrow & A \\
{\scriptstyle g}\downarrow & & \downarrow{\scriptstyle f} \\
Y & \underset{p}{\longrightarrow} & B
\end{array}
$$

is said to be a *quasi-pullback* if the corresponding map from X to $Y \times_B A$ is surjective. A quasi-pullback square in which p is surjective is said to be a *covering square*. In this case, we say that f is covered by g.

Definition 4.4. A class $\mathcal{S}$ of stable maps is said to be *representable* if there exists a function $\pi : E \to U$ for which $\mathcal{S}(\pi)$ holds and such that every function $f : A \to B$ satisfying $\mathcal{S}(f)$ is covered by a pullback of π. More explicitly, the latter condition means that we can fit f in a diagram of the form

$$
\begin{array}{ccccc}
A & \xleftarrow{\ p_1\ } & X \times_U E & \longrightarrow & E \\
{\scriptstyle f}\downarrow & & \downarrow{\scriptstyle \pi_1} & & \downarrow{\scriptstyle \pi} \\
B & \xleftarrow[\ p_0\ ]{} & X & \underset{\chi}{\longrightarrow} & U
\end{array}
\tag{4.5}
$$

where the left hand square is covering and the square on the right is a pullback.

By Theorem 3.31, the class of set-fibred maps (in a universe Type) is representable (by a function from a larger universe). Moreover, we can take the left hand square to be the identity.

4.3. *The collection axiom*

Definition 4.6. A covering square

$$\begin{array}{ccc} D & \xrightarrow{q} & B \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle f} \\ C & \xrightarrow{p} & A \end{array}$$

is said to be a *collection square* if for any $a : A$, $E :$ Type and any surjective function $e : E \twoheadrightarrow \mathsf{fib}_f(a)$,

$$\exists_{(c:\mathsf{fib}_p(a))} \exists_{(t:\mathsf{fib}_g(c)\to E)} e \circ t \sim q_c;$$

where q_c is the restriction of q to $\mathsf{fib}_g(a)$.

Definition 4.7. Let $\mathcal{S}$ be a class of small maps. The *collection axiom* is the statement $\mathsf{CA}(\mathcal{S})$ that for any small map $f : A \to X$ and any surjection $p : C \to A$ from a set C there is a quasi-pullback diagram of the form

$$\begin{array}{ccccc} B & \longrightarrow & C & \xrightarrow{p} & A \\ {\scriptstyle g}\downarrow & & & & \downarrow{\scriptstyle f} \\ Y & & \longrightarrow & & X \end{array}$$

in which g is a small map and the bottom map is surjective.

Since we have an object classifier, we can replace maps with types in a context. After this transformation, the collection axiom becomes:

$$(\forall_{(a:A)} \exists_{(c:C)} R(a,c)) \to \exists_{(B:\mathcal{S})} \exists_{(f:B\to C)} \forall_{(a:A)} \exists_{(b:B)} R(a, f(b))$$

where C is a set, $A : \mathcal{S}$ and $R : A \to C \to$ Prop.

This axiom is often included in the axioms for algebraic set theory and hence in predicative topos theory. It seems unlikely that this axiom is provable in homotopy type theory, simply because none of its axioms seem applicable. In the constructive set theory CZF (Aczel and Rathjen 2001), unlike in classical Zermelo set theory, the collection axiom is *stronger* than the replacement axiom. The replacement axiom *is* derivable from the resizing rules; see Section 3.4. In line with Voevodsky's proposal to add resizing rules to homotopy type theory, one could also consider its extension with the collection axiom.

An cumulative hierarchy of sets may be defined using higher inductive types (see The Univalent Foundations Program 2013). The induced set theory does satisfy the replacement axiom.

4.4. *The axiom of multiple choice*

Definition 4.8. The AMC is the statement that any function f fits on the right in a collection square.

The AMC implies the collection axiom (Moerdijk and Palmgren 2002, Proposition 4.3). Conversely, by the existence of the object classifier, the collection axiom implies AMC. This follows from a small adaptation of van den Berg (2012, Theorem 4.3).

It seems difficult to derive AMC, or equivalently the collection axiom, in a univalent type theory, even when we add resizing rules. One possible route towards a counter model would be to construct the Kan simplicial set model in ZF (without choice) and use the fact that ZF does not prove AMC (van den Berg 2012). However, this is beyond the scope of this article.

4.5. *Projective covers*

The 0-truncated types in the cubical set model are precisely the setoids. Hence, using an extensional type theory with a propositions-as-types interpretation as meta-theory, every set has a projective cover (Van Den Berg 2009). The same holds if we have the axiom of choice in the meta-theory. Both the collection axiom (Joyal and Moerdijk 1995, 5.2) and AMC follow from this axiom (Moerdijk and Palmgren 2002). On the other hand (Rezk 2010, Proposition 11.2), the 0-truncation of a model topos of simplicial sheaves on a site is the topos of sheaves on that site. Hence, if we start with a topos without countable choice, we cannot have projective covers in the cubical sets model. This suggests investigating homotopy type theory with and without this axiom. It is argued in van den Berg (2012) that we need AMC to obtain a model theory for predicative toposes with good closure properties, e.g. closure under sheaf models. Concretely, AMC is used to show that W-types are small in sheaf models, but also to show that every internal site is presentable. It would be interesting to reconsider these issues in the presence of higher inductive types and the univalence axiom.

5. Conclusion and outlook

Our work is a contribution to the program of providing an elementary (first order) definition of an ∞-topos as conjectured models of univalent homotopy type theory with higher inductive types (Shulman 2013). One would hope that many of the constructions that apply to predicative toposes (sheaves, realizability, gluing, ...) can be extended to homotopy type theory. By showing that Sets form a predicative topos, we make a small step in this direction. Our result may be compared to e.g. Proposition 11.2 in Rezk (2010): the 0-truncation of a model topos of simplicial sheaves on a site is the topos of sheaves on that site. Shulman (2013) shows that univalence is stable under gluing.

Moreover, this research program should contribute to a better understanding of the model theory of type theory based proof assistants such as Coq (Coq Development Team 2012) and agda (Norell 2007) improving the set theoretical models; e.g. Werner (1997). These proof assistants currently lack subset types, quotient types, functional extensionality,

proof irrelevance for mere propositions, etc. In univalent homotopy type theory, we have all these features: Set forms a predicative topos with Prop as a (large) subobject classifier and the universe acts as an object classifier. This should facilitate the formalization of Mathematics in a univalent type theory, especially now that a computational interpretation of the univalent axiom has been verified in a model (Barras 2013; Bezem *et al.* 2014). These models extend the known setoid models which already model some of the features above. The semantics of small induction-recursion (Hancock *et al.* 2013) depends on the set theoretic equivalence of the functor category and the slice category. The object classifier in Section 3.5 internalizes this. Moreover, it captures a similar kind of smallness; e.g. Theorem 4.2.

Acknowledgements

References

Aczel, P. and Gambino, N. (2002) Collection principles in dependent type theory. In: *Types for Proofs and Programs*, Springer 1–23. `doi:10.1007/3-540-45842-5_1`.

Aczel, P. and Rathjen, M. (2001) Notes on constructive set theory.

Ahrens, B., Kapulkin, K. and Shulman, M. (2013) Univalent categories and the Rezk completion. *MSCS*. `arXiv:1303.0584`.

Altenkirch, T. (1999) Extensional equality in intensional type theory. In: *14th Symposium on Logic in Computer Science* 412–420. `doi:10.1109/LICS.1999.782636`.

Avigad, J., Kapulkin, K. and Lumsdaine, P. (2013) Homotopy limits in Coq. `arXiv:1304.0680`.

Awodey, S. (2012) Type theory and homotopy. *Epistemology versus Ontology* **27** 183–201.

Awodey, S. and Bauer, A. (2004) Propositions as [types]. *Journal of Logic and Computation* **14** (4) 447–471.

Awodey, S., Gambino, N. and Sojakova, K. (2012) Inductive types in homotopy type theory. In: *Logic in Computer Science* , IEEE 95–104.

Barras, B. (2013) Native implementation of higher inductive types (HITs) in Coq. *Technical report*. Available at: `http://www.crm.cat/en/Activities/Documents/barras-crm-2013.pdf`.

Barras, B., Coquand, T. and Huber, S. (2013) A generalization of Takeuti-Gandy interpretation. Available at: `http://uf-ias-2012.wikispaces.com/file/view/semi.pdf`.

Bezem, M., Coquand, T. and Huber, S. (2014) A model of type theory in cubical sets. *19th International Conference on Types for Proofs and Programs (TYPES 2013)*, LIPIcs, volume 26, 107–128. `doi:10.4230/LIPIcs.TYPES.2013.107`.

Bishop, E. (1967) *Foundations of Constructive Analysis*, McGraw-Hill.

Coq Development Team. (2012) *The Coq Proof Assistant Reference Manual*, INRIA-Rocquencourt.

Danielsson, N. A. (2013) Closure properties for h-levels. *Technical report*, Available at: `http://www.cse.chalmers.se/~nad/listings/equality/H-level.Closure.html`.

Hancock, P., McBride, C., Ghani, N., Malatesta, L. and Altenkirch, T. (2013) Small induction recursion. In: Hasegawa, M. (ed.) *Typed Lambda Calculi and Applications*, Lecture Notes in Computer Science volume 7941, Springer 156–172. `doi:10.1007/978-3-642-38946-7_13`.

Hofmann, M. (1995) *Extensional Concepts in Intensional Type Theory*, Ph.D. thesis, University of Edinburgh.

Hofmann, M. and Streicher, T. (1998) The groupoid interpretation of type theory. In: *Twenty-Five Years of Constructive Type Theory (Venice, 1995)* **36** 83–111.

Jacobs, B. (1999) *Categorical Logic and Type Theory*, volume 141, Elsevier.

Johnstone, P. (2002) *Sketches of an Elephant: A Topos Theory Compendium*, Oxford.

Joyal, A. and Moerdijk, I. (1995) *Algebraic Set Theory*, London Mathematical Society Lecture Note Series volume 220, Cambridge. `doi:10.1017/CBO9780511752483`.

Kapulkin, K., Lumsdaine, P. and Voevodsky, V. (2012) The simplicial model of univalent foundations. `arXiv:1211.2851`.

Lambek, J. and Scott, P. (1988) *Introduction to Higher-Order Categorical Logic*, volume 7, Cambridge University Press.

Lawvere, F. and Rosebrugh, R. (2003) *Sets for Mathematics*, Cambridge. `doi:10.1017/CBO9780511755460`.

Lumsdaine, P. and Shulman, M. (2015) Higher inductive types. In preparation.

Lurie, J. (2009) *Higher Topos Theory*, volume 170, Princeton.

Mac Lane, S. and Moerdijk, I. Topos theory. In: *Handbook of Algebra*, volume 1, Elsevier 501–528. `doi:10.1016/S1570-7954(96)80018-0`.

Maietti, M. and Sambin, G. (2005) Toward a minimalist foundation for constructive mathematics. *From Sets and Types to Topology and Analysis: Practicable Foundations for Constructive Mathematics* **48** 91–114.

Maietti, M. E. (2005) Modular correspondence between dependent type theories and categories including pretopoi and topoi. *Mathematical Structures in Computer Science* **15** 1089–1149, . Available at: `http://journals.cambridge.org/article_S0960129505004962`, `doi:10.1017/S0960129505004962`.

Maietti, M. E. and Rosolini, G. (2012) Elementary quotient completion. *ArXiv:1206.0162*.

Mines, R., Richman, F. and Ruitenburg, W. (1988) *A Course in Constructive Algebra*, Springer. `doi:10.1007/978-1-4419-8640-5`.

Moerdijk, I. and Palmgren, E. (2002) Type theories, toposes and constructive set theory: Predicative aspects of AST. *Annals of Pure and Applied Logic* **114** (1-3) 155–201. `doi:10.1016/S0168-0072(01)00079-3`.

Norell, U. (2007) *Towards a Practical Programming Language Based on Dependent Type Theory*, Chalmers University of Technology. Ph.D. thesis, Chalmers University of Technology.

Palmgren, E. (2012) Constructivist and structuralist foundations: Bishops and Lawveres theories of sets. *Annals of Pure and Applied Logic*. `doi:10.1016/j.apal.2012.01.011`.

Pelayo, Á., Voevodsky, V. and Warren, M. (2013) A preliminary univalent formalization of the p-adic numbers. `arXiv:1302.1207`.

Pelayo, Á. and Warren, M. (2012) Homotopy type theory and Voevodsky's univalent foundations. `arXiv:1210.5658`.

Rezk, C. (2010) Toposes and homotopy toposes (version 0.15). Available at: `http://www.math.uiuc.edu/~rezk/homotopy-topos-sketch.pdf`.

Rijke, E. and Spitters, B. (2013a) Limits and colimits in homotopy type theory. In preparation.

Rijke, E. and Spitters, B. (2013b) Orthogonal factorization systems in homotopy type theory. *Technical report*. Available at: `http://uf-ias-2012.wikispaces.com/file/view/images.pdf/401765624/images.pdf`.

Shulman, M. (2013) The univalence axiom for inverse diagrams and homotopy canonicity. *MSCS*. `arXiv:1203.3253`.

Spiwack, A. (2011) *A Journey Exploring the Power and Limits of Dependent Type Theory*, Ph.D. thesis, École Polytechnique, Palaiseau, France.

Streicher, T. (1992) Independence of the induction principle and the axiom of choice in the pure calculus of constructions. *Theoretical Computer Science* **103** (2) 395–408. `doi:10.1016/0304-3975(92)90021-7`.

Streicher, T. (2014) A model of type theory in simplicial sets. *Journal of Applied Logic* **12** (1) 45–49. `doi:10.1016/j.jal.2013.04.001`.

The Univalent Foundations Program. (2013) *Homotopy Type Theory: Univalent Foundations of Mathematics*, Institute for Advanced Study. `arXiv:1308.0729`.

Van Den Berg, B. (2009) Three extensional models of type theory. *Mathematical Structures in Computer Science* **19** (2) 417–434. `doi:10.1017/S0960129509007440`.

van den Berg, B. (2012) Predicative toposes. `arXiv:1207.0959`.

Voevodsky, V. (2012) A universe polymorphic type system. Available at: `http://uf-ias-2012.wikispaces.com/file/view/Universe+polymorphic+type+sytem.pdf`.

Voevodsky, V. (2014) Experimental library of univalent formalization of mathematics. *Technical report*. `arXiv:1401.0053`.

Werner, B. (1997) Sets in types, types in sets. In: *Theoretical Aspects of Computer Software*, Springer 530–546. `doi:10.1016/j.jal.2013.04.001`.

Wilander, O. (2010) Setoids and universes. *Mathematical Structures in Computer Science* **20** (4) 563–576. `doi:10.1017/S0960129510000071`.