On equality of objects in categories in constructive type theory

Erik Palmgren
Department of Mathematics, Stockholm University, SE–106 91 Stockholm, Sweden
palmgren@math.su.se

Abstract

In this note we remark on the problem of equality of objects in categories formalized in Martin-Löf's constructive type theory. A standard notion of category in this system is E-category, where no such equality is specified. The main observation here is that there is no general extension of E-categories to categories with equality on objects, unless the principle Uniqueness of Identity Proofs (UIP) holds. We also introduce the notion of an H-category with equality on objects, which makes it easy to compare to the notion of univalent category proposed for Univalent Type Theory by Ahrens, Kapulkin and Shulman.

2012 ACM Subject Classification Theory of computation – Logic – Type theory

Keywords and phrases Type theory, formalization, category theory, setoids

Digital Object Identifier 10.4230/LIPIcs.TYPES.2017.8

1 Introduction

In this note we remark on the problem of equality of objects in categories formalized in Martin-Löf's constructive type theory. A common notion of category in this system is E-category [1], where no such equality is specified. The main observation here is that there is no general extension of E-categories to categories with equality on objects, unless the principle Uniqueness of Identity Proofs (UIP) holds. In fact, for every type A, there is an E-groupoid A^ι which cannot be so extended. We also introduce the notion of an H-category, a variant of category, which makes it easy to compare to the notion of *univalent category* proposed in Univalent Type Theory [9].

When formalizing mathematical structures in constructive type theory it is common to interpret the notion of set as a type together with an equivalence relation, and the notion of function between sets as a function or operation that preserves the equivalence relations. Such functions are called *extensional functions*. This way of interpreting sets was adopted in Bishop's seminal book [4] on constructive analysis from 1967. In type theory literature [3, 6, 8, 10] such sets are called *setoids*. Formally a setoid $X = (|X|, =_X, \mathrm{eq}_X)$ consists of a type $|X|$ together with a binary relation $=_X$, and a proof object eq_X witnessing $=_X$ being an equivalence relation. We usually suppress the proof object. An extensional function between setoids $f : X \to Y$ consists of a type-theoretic function $|f| : |X| \to |Y|$, and a proof that f respects the equivalence relations, i.e. $|f|(x) =_Y |f|(u)$ whenever $x =_X u$. One writes $x : X$ for $x : |X|$, and $f(x)$ for $|f|(x)$ to simplify notation. Every type A comes with a minimal equivalence relation $\mathrm{I}_A(\cdot, \cdot)$, the so-called identity type for A. We sometimes write $a \doteq b$ for $\mathrm{I}_A(a, b)$, when the type can be inferred. The principle of Uniqueness of Identity Proofs (UIP) for a type A states that

$$(\mathrm{UIP}_A) \qquad (\forall a, b : A)(\forall p, q : a \doteq b) p \doteq q$$

(using the propostions-as-types convention that $\forall$ is Π, $\exists$ is Σ etc.) This principle is not assumed in basic type theory, but can be proved for types A where $\mathrm{I}_A(\cdot, \cdot)$ is a decidable

23rd International Conference on Types for Proofs and Programs (TYPES 2017).
Editors: Andreas Abel, Fredrik Nordvall Forsberg, and Ambrus Kaposi; Article No. 8; pp. 8:1–8:7
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

relation (Hedberg's Theorem [9]). Another essential notion used in this paper is that of family of setoids indexed by a setoid. There are several choices that can be made but the one corresponding to fibers $\{f^{-1}(a)\}_{a\in A}$ of an extensional function $f : B \to A$ between setoids is the notion of a proof-irrelevant family. Let A be a setoid. A *proof-irrelevant family B of setoids over A*, assigns to each $a : |A|$, a setoid $B(a) = (|B(a)|, =_{B(a)}, eq_{B(a)})$, and to each proof object $p : a =_A b$ an extensional function $B(p) : B(a) \to B(b)$ (the *transport map* associated with p). The transport maps should satisfy the following conditions

- $B(p)(x) =_{B(a)} x$ for all $x : B(a)$ and $p : a =_A a$ (identity)
- $B(p)(x) =_{B(b)} B(q)(x)$ for all $x : B(a)$ and $p, q : a =_A b$ (proof-irrelevance)
- $B(q)(B(p)(x)) =_{B(c)} B(r)(x)$ for all $x : B(a)$ and $p : a =_A b$, $q : b =_A c$, $r : a =_A c$ (functoriality)

From these conditions follows easily that each $B(p)$ is an isomorphism which is independent of the proof object p. Hence *proof-irrelevance*. (An equivalent definition is obtained by considering A as a discrete E-category $A^{\#}$ (whose objects are elements of $|A|$ and whose hom-setoids are $\mathrm{Hom}(a, b) = (a =_A b, \sim)$ with $p \sim q$ always true) and B as a functor from this category to the E-category of setoids. This uses concepts only defined below.)

In Univalent Type Theory [9] the identity type is axiomatized so as to allow quotients, and many other constructions. This makes it possible to avoid the extra complexity of setoids and their defined equivalence relations.

These two approaches to type theory, may lead to different developments of category theory. In both cases there are notions of categories, *E-categories* and *precategories*, which are incomplete in some sense.

2 Categories in standard type theory

Categories [5] are commonly formalized in set theory in two ways, one is the essentially algebraic formulation, where objects, arrows, and composable arrows each form sets (or classes), with appropriate operations, and the other one is via objects and hom-sets (hom-classes). Set theory gives automatically a notion of equality on objects imposed by the equality of the theory. These definitions can be carried over to type theory and setoids, by taking care to make all constructions extensional.

In type theory, an *essentially algebraically presented category*, or *EA-category* for short, is formulated as follows. It consists of three setoids $\mathrm{Ob}(\mathcal{C})$, $\mathrm{Arr}(\mathcal{C})$ and $\mathrm{Cmp}(\mathcal{C})$ of objects, arrows and composable pairs of arrows, respectively. Objects are thus supposed to be equipped with equality. There are extensional functions, providing identity arrows to objects, $1 : \mathrm{Ob} \to \mathrm{Arr}$, providing domains and codomains to arrows $\mathrm{dom}, \mathrm{cod} : \mathrm{Arr} \to \mathrm{Ob}$, a composition function $\mathrm{cmp} : \mathrm{Cmp} \to \mathrm{Arr}$, and selection functions $\mathrm{fst}, \mathrm{snd} : \mathrm{Cmp} \to \mathrm{Arr}$ satisfying familiar equations, with the axiom that for a pair of arrows f, g:

$$\mathrm{cod}(g) = \mathrm{dom}(f) \iff (\exists u : \mathrm{Cmp})\, g = \mathrm{fst}(u) \wedge f = \mathrm{snd}(u)$$

In this case $\mathrm{cmp}(u)$ will be the composition $f \circ g$. See [5, 8] for axioms and details.

The hom-set formulation in type theory is the following [7]: A *hom-family presented category $\mathcal{C}$*, or just *HF-category*, consists of a setoid C of objects, and a (proof irrelevant) setoid family of homomorphisms Hom indexed by the product setoid $C \times C$. We have $1_a : \mathrm{Hom}(a, a)$, and an extensional composition $\circ_{a,b,c} : \mathrm{Hom}(b, c) \times \mathrm{Hom}(a, b) \to \mathrm{Hom}(a, c)$ satisfying

- $f \circ_{a,a,b} 1_a =_{\mathrm{Hom}(a,b)} f \quad 1_b \circ_{a,b,b} f = f$, if $f : \mathrm{Hom}(a, b)$,
- $f \circ_{a,c,d} (g \circ_{a,b,c} h) =_{\mathrm{Hom}(a,d)} (f \circ_{b,c,d} g) \circ_{a,b,d} h$, if $f : \mathrm{Hom}(c, d)$, $g : \mathrm{Hom}(b, c)$, $h : \mathrm{Hom}(a, b)$.

For $p : a =_C c$ and $q : b =_C d$, the transport map goes as follows

$$\mathrm{Hom}(p, q) : \mathrm{Hom}(a, b) \to \mathrm{Hom}(c, d).$$

The transport maps have to satisfy the following coherence conditions:

- $\mathrm{Hom}(p, p)(1_a) =_{\mathrm{Hom}(a',a')} 1_{a'}$ for $p : a =_C a'$
- $\mathrm{Hom}(p, r)(f \circ_{a,b,c} g) =_{\mathrm{Hom}(a',c')} \mathrm{Hom}(q, r)(f) \circ_{a',b',c'} \mathrm{Hom}(p, q)(g)$ for $p : a =_C a'$, $q : b =_C b'$, $r : c =_C c'$, $f : \mathrm{Hom}(b, c)$ and $g : \mathrm{Hom}(a, b)$.

▶ Remark. The coherence conditions can be captured more briefly by just stating that 1 and $\circ$ are elements in the following dependent product setoids

(a) $1 : \Pi(C, \mathrm{Hom}\langle \mathrm{id}_C, \mathrm{id}_C\rangle)$

(b) $\circ : \Pi(C^3, \mathrm{Hom}\langle \pi_2, \pi_3\rangle \times \mathrm{Hom}\langle \pi_1, \pi_2\rangle \to \mathrm{Hom}\langle \pi_1, \pi_3\rangle)$.

In more detail, the product setoids in (a) and (b) are made using the following constructions:

Let $\mathrm{Fam}(A)$ denote the type of proof irrelevant families over the setoid A. Such families are closed under the following pointwise operations:

If $F, G : \mathrm{Fam}(A)$, then $F \times G : \mathrm{Fam}(A)$ and $F \to G : \mathrm{Fam}(A)$.

If $F : \mathrm{Fam}(A)$, and $f : B \to A$ is extensional, then the composition $Ff : \mathrm{Fam}(B)$.

The cartesian product $\Pi(A, F)$ of a family $F : \mathrm{Fam}(A)$ consists of pairs $f = (|f|, \mathrm{ext}_f)$ where $f : (\Pi x : |A|)|F(x)|$ and ext_f is a proof object that witnesses that $|f|$ is extensional, that is

$$\mathrm{ext}_f : (\forall x, y : A)(\forall p : x =_A y)[F(p)(|f|(x)) =_{F(y)} |f|(y)].$$

Two such pairs f and f' are extensionally equally if and only if $|f|(x) =_{F(x)} |f'|(x)$ for all $x : A$. Then it is straightforward to check that $\Pi(A, F)$ is a setoid.

3 E-categories and H-categories in standard type theory

According to the philosophy of category theory, truly categorical notions should not refer to equality of objects. This has a very natural realization in type theory, since there, unlike in set theory, we can choose *not to impose* an equality on a type. This leads to the notion of *E-category* [1], which is essentially an HF-category with equality on objects taken away, and the corresponding transport maps removed.

An *E-category* $\mathcal{C} = (C, \mathrm{Hom}, \circ, 1)$ is the formulation of a category where there is a *type* C of objects, but no imposed equality, and for each pair of objects a, b there is a setoid $\mathrm{Hom}(a, b)$ of morphisms from a to b. The composition is an extensional function

$$\circ : \mathrm{Hom}(b, c) \times \mathrm{Hom}(a, b) \to \mathrm{Hom}_{\mathcal{C}}(a, c).$$

satisfying the familiar laws of associativity and identity. A functor or an *E-functor* between E-categories is defined as usual, but the object part does not need to respect any equality of objects (because there is none).

Now an interesting question is whether we can impose an equality of objects onto an E-category which is compatible with composition, so as to obtain an HF-category? We may consider an intermediate structure on E-categories as follows.

Define an *H-category* $\mathcal{C} = (C, =_C, \mathrm{Hom}, \circ, 1, \tau)$ to be an E-category with an equivalence relation $=_C$ on the objects C, and a family of isomorphisms $\tau_{a,b,p} \in \mathrm{Hom}(a, b)$, for each proof $p : a =_C b$. The morphisms should satisfy the conditions

(H1) $\tau_{a,a,p} = 1_a$ for any $p : a =_C a$

(H2) $\tau_{a,b,p} = \tau_{a,b,q}$ for any $p, q : a =_C b$

(H3) $\tau_{b,c,q} \circ \tau_{a,b,p} = \tau_{a,c,r}$ for any $p : a =_C b$, $q : b =_C c$ and $r : a =_C c$.

Axioms (H1) and (H3) can be replaced by the special cases $\tau_{a,a,\mathrm{ref}(a)} = 1_a$, and $\tau_{b,c,q} \circ \tau_{a,b,p} = \tau_{a,c,\mathrm{tr}(q,p)}$ where ref and tr are specific proofs of reflexivity and transitivity. Note that by these axioms, it follows that each $\tau_{a,b,p}$ is indeed an isomorphism.

A functor between H-categories $\mathcal{C} = (C, =_C, \mathrm{Hom}, \circ, 1, \tau)$ and $\mathcal{D} = (D, =_D \mathrm{Hom}', \circ', 1', \sigma)$ is an E-functor F from $(C, \mathrm{Hom}, \circ, 1)$ to $(D, \mathrm{Hom}', \circ', 1')$ such that $a =_C b$ implies $F(a) =_D F(b)$ and $F(\tau_{a,b,p}) = \sigma_{F(a),F(b),q}$ for $p : a =_C b$ and $q : F(a) =_D F(b)$.

An H-category $\mathcal{C}$ is called *skeletal* if $a =_C b$ whenever a and b are isomorphic in $\mathcal{C}$.

To pass between H- and HF-categories we proceed as follows:
For an H-category $\mathcal{C} = (C, =_C, \mathrm{Hom}, \circ, 1, \tau)$, define a transportation function

$$\mathrm{Hom}(p, q) : \mathrm{Hom}(a, b) \to \mathrm{Hom}(a', b')$$

for $p : a =_C a'$ and $q : b =_C b'$, by

$$\mathrm{Hom}(p, q)(f) = \tau_{b,b',q} \circ f \circ \tau_{a',a,p^{-1}}.$$

It is straightforward to check that this defines an HF-category.

Conversely, an HF-category $\mathcal{C} = (C, \mathrm{Hom}, \circ, 1)$ yields an E-category $(|C|, \mathrm{Hom}, \circ, 1)$ and we can define, an H-structure on it by, for $p : a =_C b$,

$$\tau_{a,b,p} = \mathrm{Hom}(r(a), p)(1_a) : \mathrm{Hom}(a, b).$$

These constructions are inverses to each other, though they do not form an equivalence, since the two categories have different notions of functors.

4 E-categories are proper generalizations of H-categories

The existence of some H-structure on any E-category turns out to be equivalent to UIP.

▶ **Theorem 1.** *If UIP holds for the type C, then any E-category with objects C can be extended to an H-category.*

Proof. The equivalence relation on C will be $\mathrm{I}_C(\cdot, \cdot)$. Using induction on identity one defines $\tau_{a,b,p} \in \mathrm{Hom}(a, b)$ for $p \in \mathrm{I}(C, a, b)$ by

$$\tau_{a,a,\mathrm{ref}(a)} =_{\mathrm{def}} \mathrm{id}_a.$$

The UIP property implies (H2). Property (H3) follows from transitivity and (H2). ◀

▶ Remark. We recall that by Hedberg's theorem, UIP holds for a type C, whenever $\mathrm{I}_C(x, y) \vee \neg \mathrm{I}_C(x, y)$, for all $x, y : C$. This explains why the extension problem is trivial in a classical setting.

Let A be an arbitrary type. Define the E-category A^ι where A is the type of objects, and hom setoids are given by

$$\mathrm{Hom}(a, b) =_{\mathrm{def}} (\mathrm{I}_A(a, b), \approx)$$

where $p \approx q$ holds if and only if $\mathrm{I}_{\mathrm{I}_A(a,b)}(p, q)$ is inhabited. Let composition be given by the proof object transitivity, and the identity on a is $\mathrm{ref}(a)$. Then it is well-known that A^ι is an E-groupoid.

▶ **Theorem 2.** *Let A be a type. Suppose that the E-category A^ι can be extended to an H-category. Then UIP holds for A.*

Proof. Suppose that $=_A, \tau$ is an H-structure on A^ι.

Now since $\mathrm{I}_A(a,b)$ is the minimal equivalence relation on A, there is a proof object $f(p) : a =_A b$ for each $p : \mathrm{I}_A(a,b)$. Thus $\tau_{a,b,f(p)} : \mathrm{Hom}(a,b) = \mathrm{I}_A(a,b)$. Let $D(a,b,p)$ be the proposition

$$\tau_{a,b,f(p)} \approx p. \tag{1}$$

By (H1) it holds that

$$\tau_{a,a,f(\mathrm{ref}(a))} \approx \mathrm{ref}(a),$$

i.e. $D(a,a,\mathrm{ref}(a))$. Hence by I-elimination (1) holds. On the other hand, (H1) gives for $p : \mathrm{I}_A(a,a)$, that

$$\tau_{a,a,f(p)} \approx \mathrm{ref}(a). \tag{2}$$

With (1) this gives

$$p \approx \mathrm{ref}(a)$$

for any $p : \mathrm{I}_A(a,a)$, which is equivalent to UIP for A. ◀

▶ **Corollary 3.** *Assuming any E-category with A as the type of objects can be extended to an H-category. Then UIP holds for A.*

In classical category theory any category may be equipped with isomorphism as equality of objects (see remark above). This is thus *not* possible in basic type theory, with the above A^ι as counter examples.

5 Categories in Univalent Type Theory

In Univalent Type Theory [9], a *set* is a type that satisfies the UIP condition. A *precategory* [9, Chapter 9.1] is a tuple $\mathcal{C} = (C, \mathrm{Hom}, \circ, 1)$ where C is a type, Hom is a family of types over $C \times C$, such that $\mathrm{Hom}(a,b)$ is a set for all $a, b : C$. Moreover $1_a : \mathrm{Hom}(a,a)$ and

$$\circ : \mathrm{Hom}(b,c) \times \mathrm{Hom}(a,b) \to \mathrm{Hom}(a,c)$$

satisfy the associativity and unit laws up to I-equality.

Such a precategory thus forms an E-category by considering the hom-set as the setoid $(\mathrm{Hom}(a,b), \mathrm{I}_{\mathrm{Hom}(a,b)}(\cdot,\cdot))$. We have moreover:

▶ **Theorem 4.** *Every precategory whose type of objects is a set is an H-category.*

Proof. Define $a \cong b$ to be the statement that a and b are isomorphic in $\mathcal{C}$ i.e.

$$(\exists f : \mathrm{Hom}(a,b))(\exists g : \mathrm{Hom}(b,a))\, g \circ f \doteq 1_a \wedge f \circ g \doteq 1_b.$$

By I-elimination one defines a function

$$\sigma_{a,b} : a \doteq b \to a \cong b \tag{3}$$

by $\sigma_{a,a}(\mathrm{ref}(a)) = (1_a, (1_a, (\mathrm{ref}(1_a), \mathrm{ref}(1_a))))$. Define by taking the first projection $\tau_{a,b,p} = (\sigma_{a,b}(p))_1 : \mathrm{Hom}(a,b)$. By I-induction it follows that

$\tau_{a,a,\mathrm{ref}(a)} \doteq 1_a$ for any $p : a \doteq a$,

$\tau_{b,c,q} \circ \tau_{a,b,p} \doteq \tau_{a,c,q\circ p}$ for any $p : a \doteq b$ and $q : b \doteq c$.

For a precategory where C is a set, it follows that for any $p, q : a \doteq b$ such that $p \doteq q$ holds, so by substitution

$$\tau_{a,b,p} = \tau_{a,b,q}.$$

Thus τ gives an H-structure on C, so the precategory is in fact an H-category. ◀

An *univalent category,* or UF-category, is a precategory where the function $\sigma_{a,b}$ in (3) is an equivalence for any $a, b : C$; see [2] and [9, Chapter 9.1]. In particular, it means that if $a \cong b$, then $\mathrm{I}_C(a, b)$.

▶ **Example 5.** An example of a precategory which is not a univalent category is given by $C = \mathrm{N}_2$ where $\mathrm{Hom}(m, n) = \mathrm{N}_1$. Here $0 \cong 1$, but $\mathrm{I}_C(0, 1)$ is false.

▶ Remark. Note that a UF-category whose type of objects is a set, is a skeletal H-category.

The reverse is however not true.

▶ **Example 6.** Suppose that $\mathcal{C}$ is a skeletal precategory whose type of objects is a set. Is $\mathcal{C}$ necessarily a univalent category? No. Consider the group $\mathbb{Z}_2$ as a one object, skeletal precategory: Let the underlying set be N_1 and $\mathrm{Hom}(0, 0) = \mathrm{N}_2$ with 0 as unit and $\circ$ as addition. This is not a univalent category, compare Example 9.15 in [9]. Thus the standard multiplication table presentation of a nontrivial group is not a univalent category.

6 Conclusion

In conclusion, the notion of univalent category is too restrictive to cover many familiar examples. H-category is generalization of precategory and is a convenient version of E-category with equality on objects. The notion of E-category is still more general as shown here.

7 Acknowledgement

The author is grateful to the referees for spotting a false theorem in a previous version of this paper.

References

1. Peter Aczel. Galois: a theory development project. Report for the 1993 Turin meeting on the Representation of Mathematics in Logical Frameworks, 1995. Available from: `http://www.cs.man.ac.uk/~petera/papers.html`.
2. Benedikt Ahrens, Krzysztof Kapulkin, and Michael Shulman. Univalent categories and the Rezk completion. *Math. Structures Comput. Sci.*, 25(5):1010–1039, 2015.
3. Gilles Barthe, Venanzio Capretta, and Olivier Pons. Setoids in type theory. *Journal of Functional Programming*, 13:261–293, 2003.
4. Errett Bishop. *Foundations of constructive analysis*. McGraw-Hill Book Co., New York-Toronto, Ont.-London, 1967.
5. Saunders Mac Lane. *Categories for the working mathematician*, volume 5 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1998.
6. Erik Palmgren. Proof-relevance of families of setoids and identity in type theory. *Arch. Math. Logic*, 51(1-2):35–47, 2012.

7 Erik Palmgren. Constructions of categories of setoids from proof-irrelevant families. *Arch. Math. Logic*, 56(1-2):51–66, 2017.

8 Erik Palmgren and Olov Wilander. Constructing categories and setoids of setoids in type theory. *Log. Methods Comput. Sci.*, 10(3):3:25, 14, 2014.

9 The Univalent Foundations Program. *Homotopy type theory—univalent foundations of mathematics*. The Univalent Foundations Program, Princeton, NJ; Institute for Advanced Study (IAS), Princeton, NJ, 2013.

10 Olov Wilander. Constructing a small category of setoids. *Math. Structures Comput. Sci.*, 22(1):103–121, 2012.