ON EQUALITY OF OBJECTS IN CATEGORIES IN CONSTRUCTIVE TYPE THEORY

ERIK PALMGREN

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, a variant of 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.

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. 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 [3] on constructive analysis from 1967. In type theory literature such sets are called *setoids*. Formally a setoid $X = (|X|, =_X)$ consists of a type $|X|$ together with a binary relation $=_X$, and a proof object for $=_X$ being an equivalence relation. An extensional function between setoids $f : X \longrightarrow Y$ consists of a type-theoretic function $|f| : |X| \longrightarrow |Y|$, 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. When the type can be inferred we write $a \doteq b$ for $\mathrm{I}_A(a, b)$. The principle of Uniqueness of Identity Proofs (UIP) for a type A states that

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

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 relation (Hedberg's Theorem [9]).

Date: August 5, 2017.

In Univalent Type Theory [9] the identity type is axiomatized so as allow to 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, 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.

1. Categories in standard type theory

A *category with equality of objects* can be formulated in an essentially algebraically manner in type theory. 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. There are extensional functions, providing identity arrows to object, $1 : \mathrm{Ob} \longrightarrow \mathrm{Arr}$, providing domains and codomains to arrowwsdom, $\mathrm{cod} : \mathrm{Arr} \longrightarrow \mathrm{Ob}$, a composition function $\mathrm{cmp} : \mathrm{Cmp} \longrightarrow \mathrm{Arr}$, and selection functions $\mathrm{fst}, \mathrm{snd} : \mathrm{Cmp} \longrightarrow \mathrm{Arr}$ satisfying familiar equations, with the convention that for a composable pair of arrows u: $\mathrm{cod}(\mathrm{fst}(u)) = \mathrm{dom}(\mathrm{snd}(u))$. See [4, 7] for details.

An equivalent 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 $C \times C$. Moreover there are elements in the following dependent product setoids

(a) $1 : \Pi(\mathrm{Ob}(\mathcal{C}), \mathrm{Hom}\langle \mathrm{id}_{\mathrm{Ob}(\mathcal{C})}, \mathrm{id}_{\mathrm{Ob}(\mathcal{C})}\rangle)$
(b) $\circ : \Pi(\mathrm{Ob}(\mathcal{C})^3, \mathrm{Hom}\langle \pi_2, \pi_3\rangle \times \mathrm{Hom}\langle \pi_1, \pi_2\rangle \longrightarrow \mathrm{Hom}\langle \pi_1, \pi_3\rangle)$.

satisfying

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

Here $g \circ_{a,b,c} h$ is notation for the application $\circ((a,b,c),(g,h))$.

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 $F \longrightarrow G : \mathrm{Fam}(A)$.

If $F : \mathrm{Fam}(A)$, and $f : B \longrightarrow A$ 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 that $|f|$ is extensional, which is stated as

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

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 easy to check that $\Pi(A, F)$ is a setoid.

2. 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*.

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

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

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 a 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.

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 morphisms $\tau_{a,b,p} \in \mathrm{Hom}(a, b)$, 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 an isomorphism. Specifying an H-structure on an E-category $\mathcal{C} = (C, \mathrm{Hom}, \circ, 1)$ then clearly amounts to providing an equivalence relation $=_C$ and an E-functor $(C, =_C)^{\#} \longrightarrow \mathcal{C}$. Here is $(C, =_C)^{\#}$ is the E-category with objects C and $(a =_C b, \sim)$ as the Hom setoid, where $p \sim q$ is always true.

An H-category $\mathcal{C}$ is *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

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

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 $p : a =_C b$,

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

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 $q : F(a) =_D F(b)$.

We consider the problem of extending a E-category to an H-category, first from the classical point of view. The non-constructive *Zermelo axiom of choice* (ZAC) may be stated as follows using setoids:

> For any setoids A and B, and any binary relation R between A and B, satisfying the totality condition $(\forall x : A)(\exists y : B)\, R(x, y)$, there is an extensional function $f : A \longrightarrow B$ such that $(\forall x : A)\, R(x, f(x))$.

The principle ZAC implies the principle of excluded middle, PEM, by Diaconescu's Theorem. Then by Hedberg's Theorem also UIP holds [9].

For any type S, let $\hat{S}$ be the setoid $(S, \mathrm{I}_S(\cdot,\cdot))$. It is well known that the special case of ZAC where A is such a setoid can be proven in constructive type theory. We call this special case the *type theoretic axiom of choice* (TTAC).

For setoid A write $\bar{A}$ for $\widehat{|A|}$.

Theorem 2.1. *(Using ZAC) For any setoid A there is an extensional function $s : A \longrightarrow \bar{A}$ such that, for all $x, u : A$, $x =_{\bar{A}} s(x)$, and*

$$x =_A u \Longrightarrow s(x) =_{\bar{A}} s(u). \tag{1}$$

Proof. Let $B = \bar{A}$. Apply ZAC to the trivially true statement

$$(\forall x : A)(\exists y : B)\mathrm{I}_{|A|}(x, y).$$

This gives the required extensional function s, and extensionality implies that (1) holds. □

The significance of this theorem is that the function s selects exactly one element from each equivalence class that $=_A$ defines, i.e. (1). Note that the existence of such selection functions, and the TTAC implies the general ZAC. We refer to [5] for the discussion of Zermelo's axiom of choice from the type-theoretic perspective.

Any E-category with an equivalence relation on objects, that refine the isomorphism relation, may be extended to an H-category using ZAC.

Theorem 2.2. *(ZAC) Assume that $\mathcal{C} = (C, \mathrm{Hom}, \circ, 1)$ is an E-category, and $=_C$ is an equivalence relation on C, such that a and b are isomorphic, whenever $a =_C b$. Then there is a τ giving an H-structure $(=_C, \tau)$ on $\mathcal{C}$.*

Proof. By the assumption, we have a proof object σ such that for each $p : a =_C b$, $\sigma_{a,b,p} : \mathrm{Hom}(a, b)$ is an isomorphism. By Theorem 2.1 there is a proof object g such that for all $a : C$

$$g(a) : \mathrm{I}_{|C|}(a, s(a)).$$

Since $\mathrm{I}_{|C|}(\cdot,\cdot)$ is the minimal equivalence relation on $|C|$ there is a proof object f such that

$$f : (\forall ab : |C|)(\mathrm{I}_{|C|}(a, b) \longrightarrow a =_C b).$$

Thereby we have for each $a : C$ an isomorphism in $\mathcal{C}$,

$$\phi_a = \sigma_{a,s(a),f(a,s(a),g(a))} : \mathrm{Hom}(a, s(a)).$$

Using induction on identity one defines $\rho_{a,b,p} : \mathrm{Hom}(a, b)$ for $p : \mathrm{I}_C(a, b)$ by

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

The UIP property implies (H2). Property (H3) follows from transitivity and (H2). Now by Theorem 2.1 there is a proof object h such that for $a, b : C$ and $p : a =_C b$,

$$h(a, b, p) : \mathrm{I}_C(s(a), s(b)).$$

Finally for $p : a =_C b$, we define the isomorphism

$$\tau_{a,b,p} = \phi_b^{-1} \circ \rho_{s(a),s(b),h(a,b,p)} \circ \phi_a.$$

By (H1) – (H3) for ρ, it follows, using the inverses, that also τ has these properties. $\square$

3. 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 3.1. *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). $\square$

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 3.2. *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

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

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 (2) holds. On the other hand, (H1) gives for $p : \mathrm{I}_A(a, a)$, that

$$\tau_{a,a,f(p)} \approx \mathrm{ref}(a). \tag{3}$$

With (2) this gives

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

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

Corollary 3.3. *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 maybe equipped with isomorphism as equality of objects (using Theorem 2.2). This is thus *not* possible in basic type theory, with the A^ι as counter examples.

4. Categories in Univalent Type Theory

In Univalent Type Theory [9], the notion of 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 any $a, b : C$. Moreover $1_a : \mathrm{Hom}(a, a)$ and

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

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 $(\mathrm{Hom}(a, b), \mathrm{I}_{\mathrm{Hom}(a,b)}(\cdot, \cdot))$.

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

$$\sigma_{a,b} : \mathrm{I}_C(a, b) \longrightarrow a \cong b \tag{4}$$

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 $p : \mathrm{I}_C(a, a)$,
$\tau_{b,c,q} \circ \tau_{a,b,p} \doteq \tau_{a,c,q\circ p}$ for any $p : \mathrm{I}_C(a, b)$ and $q : \mathrm{I}_C(b, c)$.

For a precategory where C is a set, it follows that for any $p, q : \mathrm{I}_C(a, b)$, $\mathrm{I}_{\mathrm{I}_C(a,b)}(p, q)$ 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* is a precategory where the function $\sigma_{a,b}$ in (4) is an equivalence for any $a, b : C$; see [1] and [9, Chapter 9.1]. In particular, it means that if $a \cong b$, then $\mathrm{I}_C(a, b)$.

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.

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

Suppose that $\mathcal{C}$ is a skeletal precategory whose type of objects is a set. Is $\mathcal{C}$ necessarily a univalent category? 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.

References

[1] Benedikt Ahrens, Chris Kapulkin and Mike Shulman. Univalent categories and the Rezk completion. *Mathematical Structures Computer Science* 25 (2015), pp. 1010 – 1039.

[2] Gilles Barthe, Venanzio Capretta and Olivier Pons. Setoids in type theory. *Journal of Functional Programming* 13(2003), 261 – 293.

[3] Errett Bishop. *Foundations of Constructive Analysis.* McGraw-Hill 1967.

[4] Saunders MacLane. *Categories for the Working Mathematician.* Second edition. Springer 1997.

[5] Per Martin-Löf. 100 years of Zermelo's axiom of choice: what was the problem with it? *The Computer Journal* 49(2006), pp. 345 – 350.

[6] Erik Palmgren. Proof-relevance of families of setoids and identity in type theory. *Archive for Mathematical Logic* 51(2012), pp. 35 – 47.

[7] Erik Palmgren. Constructions of categories of setoids from proof-irrelevant families. *Archive for Mathematical Logic* (2017).

[8] Erik Palmgren and Olov Wilander. Constructing categories and setoids of setoids in type theory. *Logical Methods in Computer Science* 10(2014), Issue 3, paper 25.

[9] *Homotopy Type Theory: Univalent Foundations of Mathematics.* The Univalent Foundations Program, Institute for Advanced Study, Princeton 2013.

[10] Olov Wilander. Constructing a small category of setoids. *Mathematical Structures in Computer Science* 22(2012), pp. 103 – 121.

Department of Mathematics, Stockholm University, SE–106 91 Stockholm, Sweden
E-mail address: palmgren@math.su.se
URL: www.math.su.se/∼palmgren