Identity, Indiscernibility, and Univalence

A Higher Structure Identity Principle

Mike Shulman

(jww Benedikt Ahrens, Paige North, and Dimitris Tsementzis)

May 1, 2020

Contents

Contents

Question: When are two things the same?
Answer (Leibniz): When they have all the same properties.

This splits in two:

1) Indiscernibility of identicals : If $x=y$, then $P(x) \Leftrightarrow P(y)$ for all properties P.

In mathematics, usually asserted explicitly in foundational theories.

- In first order logic / ZFC, this is the substitutive principle of equality
- We will use Dependent Type Theory (DTT) where it is the eliminator of identity types:

$$\frac{x:A,\, y:A,\, p:x=y \vdash C(x,y,p)\ \mathrm{Type} \qquad x:A \vdash c(x) : C(x,x,\mathrm{refl}_x)}{x:A,\, y:A,\, p:x=y \vdash J(C,c,x,y,p) : C(x,y,p)}$$

2) Identity of indiscernibles : If $P(x) \Leftrightarrow P(y)$ for all properties P, then $x=y$.

In mathematics, usually trivially true, because of haecceities: Take $P(u) :\equiv (x=u)$.

Often, in mathematics, we want to consider things "the same" under weaker conditions:

- Isomorphism of groups, rings, fields, topological spaces, ...
- Equivalence of categories
- Biequivalence of bicategories

...

In ZFC set theory, isomorphic groups are not indiscernible! Take $P(G) := (\emptyset \in G)$.

But such properties are "uninteresting", or more precisely non-structural.

In Dependent Type Theory, we cannot exhibit any non-structural property.

This metatheoretic property shows that DTT matches mathematical practice better than ZFC, in this regard. But for practical use, we would rather have a positive, internal version.

In Homotopy Type Theory / Univalent Foundations (HoTT/UF) — a form of DTT — we can prove, internally, that all properties are structural, i.e. that isomorphic structures are indiscernible.

$\longrightarrow$ "Indiscernibility of isomorphs"

Indiscernibility of Isomorphs = Structure Identity Principle

Again because of haecceities, indiscernibility of isomorphs implies identity of isomorphs:

$$(G \cong H) \longrightarrow (G = H)$$

To a set-theoretic mind this looks like a "skeletality" property, collapsing isomorphism to identity. But actually, in HoTT/UF we expand the notion of identity to include isomorphism.

To be precise, from indiscernibility of identicals we get a canonical map in the other direction:

$$(G = H) \longrightarrow \mathrm{Iso}(G, H)$$

and we prove a Structure Identity Principle (SIP) saying that this map is an equivalence.
Hence, in particular, there is a map in the other direction; but rather more is true.

In fact, in practice we prove the SIP first and deduce indiscernibility of isomorphs from that.

The basic principle that implies all SIPs is Voevodsky's univalence axiom, which is just the SIP for types:

> For $A, B : \mathrm{Type}$, the canonical map
> $$(A = B) \to (A \simeq B)$$
> is an equivalence.

$A \simeq B$ denotes the type of equivalences. This is roughly like the type of isomorphisms (functions $f : A \to B$ and $g : B \to A$ such that $f \circ g = id_B$ and $g \circ f = id_A$) but enhanced to be correct for types that are not sets, i.e. whose identity types $x =_A y$, $u =_B v$ are not propositions (i.e. may have more than one distinct element).

Univalence (the SIP for types) implies the SIP for groups:

$$\mathrm{Group} := \sum_{G:\mathrm{Set}} \sum_{m:G\to G\to G} \sum_{e:G} \left(\prod_{a:A} m(a,e)=a \right) \times \cdots$$

Since equality of tuples is componentwise, for groups $(G,m,e,\ldots)$ and $(H,m',e',\ldots)$

we have $$\left[(G,m,e,\ldots) =_{\mathrm{Group}} (H,m',e',\ldots)\right] \simeq (G=H) \times (m=m') \times (e=e')$$

$$\simeq (G\simeq H) \times \cdots \qquad \text{(by univalence axiom)}$$

$$\simeq \left[(G,m,e,\ldots) \cong (H,m',e')\right]$$

(Technically, the products on the RHS are Sigma-types, i.e. the multiplications are identified modulo the bijection $G\simeq H$, as we expect for a group isomorphism.)

General SIPs were proven in the HoTT Book (Chapter 9) and by Coquand-Danielsson.
They apply to all "set-level structures" like groups, rings, fields, topological spaces, etc.

Contents

In category theory, isomorphism is not the "correct" notion of "sameness" for categories.

A SIP for categories was also proven in the HoTT Book and Ahrens-Kapulkin-Shulman.
It says that the canonical map

$$(\mathcal{C} = \mathcal{D}) \longrightarrow (\mathcal{C} \simeq \mathcal{D})$$

is an equivalence (of types), where $\mathcal{C} \simeq \mathcal{D}$ is the type of equivalences of categories.

But to make this true, we have to define "category" carefully...

Naive definition: A "category" in DTT consists of

- A type of objects $\mathcal{C}_0$
- A family of hom-types $\mathcal{C}(x,y)$ for $x,y:\mathcal{C}_0$
- Composition functions $\circ : \mathcal{C}(y,z) \to \mathcal{C}(x,y) \to \mathcal{C}(x,z)$
- Identities $1_x : \mathcal{C}(x,x)$
- Associativity and unit axioms $\prod_{f,g,h} h\circ(g\circ f) = (h\circ g)\circ f$, $\prod_f (f\circ 1 = f)\times(1\circ f = f)$.

If the types are arbitrary, then the "axioms" should have an infinite tower of higher coherences (an $(\infty,1)$-category). For a theory that better matches traditional 1-category theory, we require all the hom-types $\mathcal{C}(x,y)$ to be sets.

In the HoTT Book, this notion (with that restriction) is called a precategory.

What restriction should we place on the type of objects $\mathcal{C}_0$? There are three natural choices.

1) It is also a set. We call this a strict category.

- Strict category theory looks the most like classical 1-category theory.
- But the SIP is false for strict categories.
- Also, most naturally-occurring categories (e.g. Set) are not strict.

2) No restriction, i.e. work with precategories.

- The most general choice, and suffices for a good deal of category theory.
- But parts of classical category theory fail, e.g. "ff+eso implies equivalence" is false.
- The SIP is also false for general precategories.

3) Require the objects to satisfy their own SIP: the map $(x =_{\mathcal{C}_0} y) \to (x \cong y)$ is an equivalence.

We call these (univalent) categories.

- Most naturally-ocurring categories (e.g. Set, Grp, Top, ...) are univalent.
- "ff+eso implies equivalence" is true (doesn't even require AC!)
- The SIP holds.

What about other (higher-)categorical structures?

∞-categories

$\vdots$

locally cubical bicategories

enriched categories

triple categories

3-categories

monoidal categories

double categories

2-categories

fibred categories

categories

$\dagger$-categories

We want to prove a general Structure Identity Principle for higher-categorical structures.

Contents

The "univalence" condition in a univalent category may seem like a "weird and new thing" relative both to classical category theory and to set-level mathematics in HoTT/UF.

(I believe this is part of what leads some people to instead consider precategories to be the "basic notion of category" in HoTT/UF.)

I aim to convince you that it is neither weird nor new, but just a new manifestation of something familiar.

"Negative thinking": Often weird things about higher categories make more sense when we compare them to their "shadows" for "lower categories", i.e. sets, posets, truth values, etc.

Do set-level structures have a version of the "univalence" condition (a.k.a. "SIP for objects")?

Some (set-level) structures have axioms that involve equality.
Hence, we can't expect any notion of "sameness" for their elements weaker than equality.

- algebraic structures like groups, rings, fields, ...
- ordered structures (with antisymmetry) like posets, lattices, total orders, ...

Other structures don't refer to equality in their axioms. In this case, there is generally an induced "indistinguishability" relation on their elements, which may be weaker than equality.

- preorders: $x \approx y$ if $(x \leq y \wedge y \leq x)$

  X is "univalent" $(x = y \leftrightarrow x \approx y)$ iff it is a partial order.

- topological spaces: $x \approx y$ if $\forall U \in \mathcal{O}(X), (x \in U \leftrightarrow y \in U)$

  X is "univalent" $(x = y \leftrightarrow x \approx y)$ iff it is T_0.

This can be captured formally by using logic with equality or logic without equality.

E.g. in (multi-sorted, first-order, relational) logic without equality, we have

- A collection of sorts.
- A collection of relation symbols, each assigned some family of sorts as arity.
- A collection of axioms, which are formulas built from the relation symbols as atomic.

In logic with equality, each sort A is additionally equipped with a specified binary relation E_A.
This allows stating axioms involving equalities of elements, e.g. associativity in a group:

$$\forall x,y,z : G,\quad E_G\big(m(m(x,y),z),\ m(x,m(y,z))\big).$$

We additionally assume that the E relations are congruences for all the others.

Nontrivial notions of "indistinguishability" arise precisely in theories that can be formulated in logic without equality.

A theory with equality is just a special case of a theory without equality...

... except that we usually consider only models of it that are standard, meaning that the interpretation of the E relations is "real" equality in the model.

But this is just another "univalence" condition, with E playing the role of "indistinguishability"!

$$(x =_A y) \longleftrightarrow E_A(x,y)$$

A "univalence" condition for structures is already present in classical first-order logic.

We don't notice it because

- For logic with equality, it coincides with "standardness", which is so "standard" that we don't even think of it as an assumption.
- For logic without equality, it generally yields "very weak separation axioms" that we often assume anyway when restricting to "nice objects".

Contents

A set-level theory has only two levels of dependency:

0) sorts are the bottom level: rank 0.
1) relation symbols (predicates) are rank 1, depending on sorts (their arities).

(In DTT and HoTT/UF, propositions are types, predicates are dependent types.)

A higher theory has more levels of dependency. For instance, a category has

0) the sort $\mathcal{C}_0$ of objects at rank 0.

1) the sorts $\mathcal{C}(x,y)$ of morphisms, depending on the objects, at rank 1.

2) relations, such as equality of morphisms, at rank 2.

Thus, a "higher signature" has to incorporate the "dependency structure".

For simplicity, consider only relational signatures (no operations, only predicates).
It is always possible to encode operations as relations via their graphs.

Makkai used inverse categories as signatures for First-Order Logic with Dependent Sorts (FOLDS).

The signature for groups:

$$\begin{array}{ccc} T & E_G & I \\ \searrow\searrow\searrow & \downarrow\downarrow & \swarrow \\ & G & \end{array}$$

$E_G(x,y)$ means $x=y$

$T(x,y,z)$ means $x \cdot y = z$

$I(x)$ means $x=e$

The signature for categories:

$$\begin{array}{ccc} T & E_A & I \\ \searrow\searrow\searrow & \downarrow\downarrow & \swarrow \\ & A & \\ & \downarrow\downarrow & \\ & O & \end{array}$$

O = objects

$A(x,y)$ = morphisms

$E_{A,x,y}(f,g)$ means $f=g$

$T_{x,y,z}(f,g,h)$ means $g \circ f = h$

$I_x(f)$ means $f = 1_x$

(Our notion of signature is somewhat more general, but I'll ignore that today.)

$$
\begin{tikzcd}
T \arrow[dr, shift left=2] \arrow[dr] \arrow[dr, shift right=2] & E_A \arrow[d, shift left] \arrow[d, shift right] & I \arrow[dl] \\
& A \arrow[d, shift right, "s"'] \arrow[d, shift left, "t"] & \\
& O &
\end{tikzcd}
$$

The nonidentity arrows with a given sort as domain represent "all the dependencies" of that sort.

T has 3 arrows $d_0, d_1, d_2 : T \overset{\to}{\underset{\to}{\to}} A$,

and three different composites

$$sd_2 = sd_1, \quad td_2 = sd_0, \quad td_0 = td_1 \quad : T \overset{\to}{\underset{\to}{\to}} O$$

Thus, in a model, the sort T is interpreted by a family of types with 6 dependencies:

$$\prod_{x,y,z:O} A(x,y) \to A(y,z) \to A(x,z) \to \mathrm{Type}$$

We think of $T_{x,y,z}(f,g,h)$ as asserting that a triangle commutes:

$$
\begin{tikzcd}
& y \arrow[dr, "g"] \arrow[d, phantom, "\circ"] & \\
x \arrow[ur, "f"] \arrow[rr, "h"'] & {} & z
\end{tikzcd}
$$

Contents

Recall the notion of identity of indiscernibles:

If $P(x) \Leftrightarrow P(y)$ for all properties P, then $x = y$.

- As a "global" statement about a foundational system, this is trivial because of haecceities.

  But as a "local" statement about a structure for a particular signature, with P ranging only over predicates in that signature, it is nontrivial.

- Our notion of "isomorphism" or "indistinguishability" between elements of a structure will assert that they "share all the same properties" expressible in that signature.

  Thus, we call it indiscernibility, writing $x \asymp y$ for "x and y are indiscernible".

- Of course, in a higher structure, $x \asymp y$ will be a type, not just a proposition.

  We call its elements indiscernibilities.

- A structure will then be univalent if each map $(x = y) \to (x \asymp y)$ is an equivalence:

  a strong form of the "local" identity of indiscernibles.

For any signature $\mathcal{L}$, an $\mathcal{L}$-structure M can be decomposed into

1) A type family $M_\perp : \mathcal{L}_\perp \to \mathrm{Type}$, consisting of the rank-0 sorts.
2) A derivative M', which is an $\mathcal{L}'_{M_\perp}$-structure for a derived signature $\mathcal{L}'_{M_\perp}$.

<u>Example:</u> $\mathcal{L}$ = the signature for categories
$$\begin{tikzcd} T \arrow[dr, Rrightarrow] & E \arrow[d, shift left] \arrow[d, shift right] & I \arrow[dl] \\ & A \arrow[d, shift left] \arrow[d, shift right] & \\ & O & \end{tikzcd}$$
,

$M_\perp$ = the type of objects M_0.

$\mathcal{L}'_{M_0}$ = the signature for categories with fixed object type M_0.

- rank-0 sorts "$A(x,y)$" for all $x,y : M_0$ (NB: infinitely many sorts!)
- rank-1 sorts "$T_{x,y,z}$" etc., with dependencies

$$\begin{tikzcd} & T_{x,y,z} \arrow[dl] \arrow[d] \arrow[dr] & \\ A(x,y) & A(y,z) & A(x,z). \end{tikzcd}$$

(Note: ranks in $\mathcal{L}'_{M_0}$ are reduced by 1 from $\mathcal{L}$.)

For any signature $\mathcal{L}$, an $\mathcal{L}$-structure M can be decomposed into

1) A type family $M_\perp : \mathcal{L}_\perp \to \mathrm{Type}$, consisting of the rank-0 sorts.
2) A derivative M', which is an $\mathcal{L}'_{M_\perp}$-structure for a derived signature $\mathcal{L}'_{M_\perp}$.

<u>Example:</u> $\mathcal{L}$ = the signature for categories
$$\begin{array}{ccc} T & E & I \\ & A & \\ & \| & \\ & O & \end{array}$$
,

$M_\perp$ = the type of objects M_0.

$\mathcal{L}'_{M_0}$ = the signature for categories with fixed object type M_0.

$(M')_\perp$ = the types of arrows $M_A(x,y)$ for all $x,y : M_0$

$(M'')_\perp$ = the propositions $T_{x,y,z}(f,g,h)$, $E_{x,y}(f,g)$, $I_x(f)$.

M''' = empty.

Given an $\mathcal{L}$-structure M, a rank-0 sort K, and $a,b : M_K$

$M_\perp + [K]$ = $M_\perp$ with one extra element of M_K — *the joker*, $\star$.

$$M_\perp + [K] \overset{[id,a]}{\underset{[id,b]}{\rightrightarrows}} M_\perp$$ sending the joker to $a : M_K$ and $b : M_K$ resp.

<u>Definition</u> An *indiscernibility* $a \asymp b$ is an isomorphism

$$[id,a]^* M' \cong [id,b]^* M'$$ of $\mathcal{L}'_{M_\perp+[K]}$-structures

that restricts along $inl : M_\perp \hookrightarrow M_\perp + [K]$ to $id_{M'}$.

<u>Definition</u> M is *univalent at K* if $(a=b) \to (a \asymp b)$ is an equivalence for all $a,b : M_K$.

Example $\mathcal{L}$ = the signature for categories. $a, b : M_0$ objects.

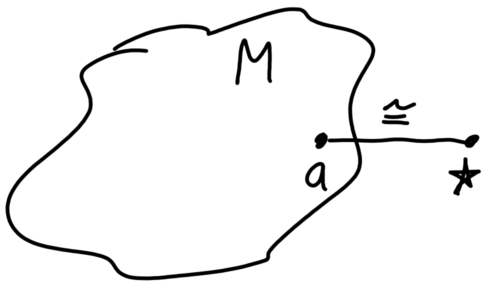
$\mathcal{L}'_{M_0 + [0]}$ = the signature for categories with objects $M_0 + 1$

$[id, a]^* M'$ = such a category with $\hom(x, \star) = \hom_M(x, a)$

$\hom(\star, y) = \hom_M(a, y)$

$\hom(\star, \star) = \hom_M(a, a)$.

= the category M with a "whisker" attached at a



An indiscernibility $\phi : a \asymp b$ consists of bijections

$\phi_{x\bullet} : \hom_M(x,a) \cong \hom_M(x,b)$ for all $x : M_0$

$\phi_{\bullet y} : \hom_M(a,y) \cong \hom_M(b,y)$ for all $y : M_0$

$\phi_{\bullet\bullet} : \hom_M(a,a) \cong \hom_M(b,b)$

(and also $\hom_M(x,y) \cong \hom_M(x,y)$ that are fixed as the identity)

that respect composition and identities in "all possible ways."

This turns out to mean

1) $\phi_{x\bullet}$ is natural in x, hence by Yoneda just an isomorphism $a \cong b$.
2) $\phi_{\bullet y}$ is natural in y, hence by Yoneda just an isomorphism $a \cong b$.
3) These two $a \cong b$ are the same, and also induce $\phi_{\bullet\bullet}$.

Thus $(a \asymp b) \simeq (a \cong b)$. So univalence at O recovers "univalent categories".

More examples

- If a sort K has an equality relation (local haecceity), then $(a \simeq b) \simeq E_K(a,b)$

  Thus, univalence at K means that M_K is a set and its equality is standard.

  This includes all set-level structures: groups, rings, fields, ...

- In particular, for morphisms in a category $f,g : M_A(x,y)$ we have $(f \simeq g) \simeq E_A(f,g)$.

  Thus, univalence at A means that M is a precategory, with standard equality.

- If P is a top-rank sort (i.e. a relation symbol), then $(a \simeq b) \simeq *$ is contractible.

  Thus, univalence at P means that M_P is a proposition — we even get <u>that</u> automatically!

  (Applies in particular to T, E, I in a category.)

Contents

Goal: $(M = N) \to (M \simeq N)$ is an equivalence for univalent $\mathcal{L}$-structures M, N.

$M \simeq N$ is the type of "equivalences of $\mathcal{L}$-structures"... but what are those?

Def 1: An equivalence of categories consists of functors $F : \mathcal{C} \to \mathcal{D}$, $G : \mathcal{D} \to \mathcal{C}$ and natural isomorphisms $FG \cong Id_{\mathcal{D}}$, $GF \cong Id_{\mathcal{C}}$.

Not clear how to generalize this to $\mathcal{L}$-structures: what is a "natural isomorphism"?

Def 2: An equivalence of categories is a functor $F : \mathcal{C} \to \mathcal{D}$ such that

1) F is essentially surjective: For all $z : \mathcal{D}_0$, there is an $x : \mathcal{C}_0$ and $Fx \cong z$.
2) F is full: For all $g : \mathcal{D}_A(Fx, Fy)$, there is an $f : \mathcal{C}_A(x, y)$ and $Ff = g$.
3) F is faithful: For all $\_ : Ff = Fg$, there is $\_ : f = g$ and $\_$.

Analogous "faithfulness" properties for the sorts T and I follow automatically.

Def: A morphism of $\mathcal{L}$-structures $F: M \to N$ is an equivalence if
For each sort K, of any rank, and each $y : N_K(F(u_1), F(u_2), \dots)$ there is a (specified) $x : M_K(u_1, u_2, \dots)$ and an indiscenibility $Fx \simeq y$.

(In N. It's a bit subtle to formulate the induction to get this right)

Write $M \simeq N$ for the type of equivalences.

Theorem (HSIP) For univalent $\mathcal{L}$-structures M, N, the map

$$(M = N) \longrightarrow (M \simeq N)$$

is an equivalence (of types).

Benedikt Ahrens, Paige Randall North, Michael Shulman, Dimitris Tsementzis.
A Higher Structure Identity Principle. arXiv:2004.06572

Thank You!