Equivalence in Foundations

Laney Gold-Rappe and Hans Halvorson

June 28, 2024

The old consensus

- Early 20th century: Zermelo-Frankel set theory won the battle about the foundations of mathematics.
- Defeated competitors:
  - Logicism
  - Finitism
  - Intuitionism
  - Type theory
- Philosophers take set theory as background framework for their inquiries.

  See: David K Lewis et al. (1986). *On the plurality of worlds*. Blackwell Oxford

New developments

- Category theory and topos theory have proved fruitful in various branches of pure mathematics (Grothendieck, Mac Lane, Lawvere)
- Martin-Löf type theory
- Computation
- Homotopy type theory (HoTT)
- Philosophical worries about set theory (structuralism, etc.)

Motivation

What got me (HH) worried about set-theoretic foundations is that it seems to have too much "sand" for use in the foundations of physics.

Isomorphic models can have different features qua sets:

- $M \vDash \phi(a)$ but $N \vDash \neg\phi(a)$.
- $\{\emptyset\} \in M$ but $\{\emptyset\} \notin N$.

Motivation

- It would seem natural to replace ZF with an elementary topos $\mathcal{E}$, thereby ignoring the (otiose?) universal membership relation.
- In a topos, $x = y$ has no meaning for $x : A$ and $y : B$, with $A \neq B$.
- Question: Can isomorphic $\mathcal{E}$-models of T have different properties? It's important to be clear here about what "different properties" means.

Motivation

- Topos-theoretic foundations are (apparently) more general. E.g. the axioms of synthetic differential geometry have no model in Sets, but do have a model in a topos.

Is a new battle coming?

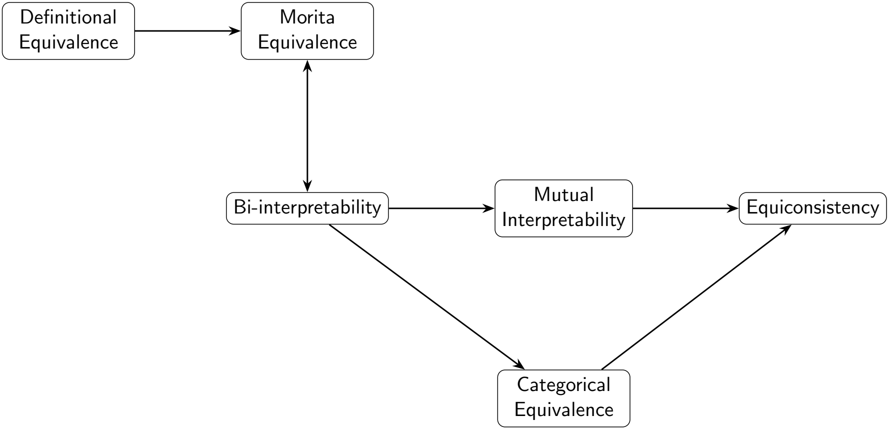
- Feferman (1969; 1977) argues against category-theoretic foundations for philosophical reasons: the idea of "aggregating" is presupposed even in category theory.
- The idea that Sets and Cats are incommensurable foundations was challenged via results of Mitchell, Osius, and Mathias
    - What exactly did they prove?

Goals of this project

- Evaluate Awodey's (2009) claim that Sets, Cats, and Types are equivalent foundations.[1]
- Evaluate whether Shulman (2019) has established that ZFC and ETCS+R are bi-interpretable.
- Sharpen the definition of bi-interpretability and compare it to other notions of equivalence.

[1] Linnebo and Rayo (2012) also suggest that Sets and Types are equivalent foundations, but don't suggest any kind of formal proof.

What do we mean by equivalent?



Equivalence and syntactic categories

- Morita equivalence (Barrett and Halvorson, 2016) is an attempt to give an elementary expression of the idea that $\mathrm{Sh}(C_T)$ and $\mathrm{Sh}(C_{T'})$ are equivalent toposes (see Makkai and Reyes, 2006).
- Morita equivalence is very likely (perhaps some fine-tuning needed?) the same thing as bi-interpretability (see Halvorson, 2019).
- Note: Morita equivalence is weaker (more liberal) than the notion that C_T and $C_{T'}$ are equivalent categories.

Why bi-interpretability matters

Informal Hypothesis: *Bi-interpretability ensures that the theories share all relevant properties.*

- Fact: Mutual interpretability does not imply bi-interpretability.
  - ZF and ZFC are mutually interpretable, but not bi-interpretable (see Enayat).
  - Hajnal Andréka, Judit Madarasz, and István Németi (1994). "Mutual definability does not imply bi-interpretability". In: *Studia Logica* 53.3, pp. 353–378. DOI: 10.1002/malq.200410051
- To do: Examples of mutually interpretable theories that have different properties (model-theoretic, proof-theoretic, etc.)

Properties preserved under equivalence

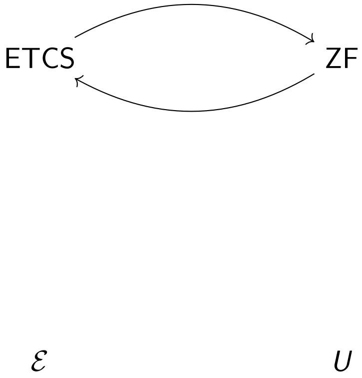
	mutual int	bi-int
κ-categorical		✓
finitely axiomatizable		✓
model complete		✓
ω-stable		
has a prime model		
strongly minimal		

Why bi-interpretability matters

Informal Hypothesis: *Bi-interpretability is our best account of expressive equivalence.*

For each Σ_1-formula ϕ, there is a Σ_2-formula $F(\phi)$ that "says the same thing".

Bi-interpretability: syntax and semantics



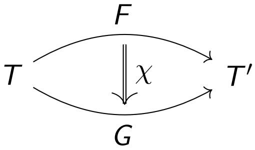
Translation

The notion of **translation** is still a work in progress[2] — although it seems to be implicit in much of the standard literature.

- A translation F has an arity n_F, which says how many variables to split a single variable into.
- A translation F has a domain formulas δ_σ^F in the target language.
- A translation represents equality $=$ in Σ in terms of some T'-provable equivalence relation in Σ'.

[2] The first explicit definition of a translation with arity $n \geq 1$ seems to be (Szczerba, 1977). The idea is developed further in (Van Benthem and Pearce, 1984; Visser, 2006). For an attempt to systematize, see (Halvorson, 2019).

2-cells between translations



Roughly speaking: A 2-cell $\chi : F \Rightarrow G$ is a formula in Σ' that represents a functional relation from the domain formula of F to the domain formula of G and that maps the extension of $F(R)$ to the extension of $G(R)$, for each relation symbol R of Σ.[3]

[3] This idea is implicit throughout model theory, and is more explicit in (Visser, 2006).

Definition

Let $F : T \to T'$ and $G : T' \to T$ be translations. We say that F and G form an **equivalence** just in case there are invertible 2-cells $\eta : 1_T \Rightarrow GF$ and $\varepsilon : 1_{T'} \Rightarrow FG$.

A translation $F : T \to T'$ determines a functor $F^* : \mathrm{Mod}(T') \to \mathrm{Mod}(T)$. See (Gajda, Krynicki, and Szczerba, 1987; Halvorson, 2019)

In particular, $F^*(M)$ is n_F copies of $D(M)$, quotiented by the equivalence relation $=_F$.

To be checked: a 2-cell $\chi : F \Rightarrow G$ should determine a natural transformation $\chi^* : F^* \Rightarrow G^*$.

Note: χ^* is not just any natural transformation, but is induced uniformly via a Σ'-formula that is a T'-provable functional relation.

Proving equivalence semantically

Given functors $f : \mathrm{Mod}(T') \to \mathrm{Mod}(T)$ and $g : \mathrm{Mod}(T) \to \mathrm{Mod}(T')$, under what conditions on f and g establish that T and T' are bi-interpretable?

See (Gajda, Krynicki, and Szczerba, 1987)

What are the natural isomorphisms on the two sides?
If ZF and ETCS are bi-interpretable, then there are linking formulas

Topos-theoretic foundations of mathematics

Definition

An **elementary topos** $\mathcal{E}$ is a category that has the following properties:

- Finite limits.
- Exponentials: For any objects $A, B \in \mathcal{E}$, there exists an object B^A and an evaluation map $ev : B^A \times A \to B$ such that for any object C and any map $f : C \times A \to B$, there is a unique map $\lambda f : C \to B^A$ making the appropriate diagram commute.
- A subobject classifier Ω: An object Ω with a morphism $true : 1 \to \Omega$ such that for any monomorphism $m : A \to B$, there exists a unique characteristic morphism $\chi_m : B \to \Omega$ making the diagram commute.

Category Axioms

Objects and Morphisms

- Two sorts: **Objects** and **Morphisms**.
- Each morphism f has a **domain** $\mathrm{dom}(f)$ and **codomain** $\mathrm{cod}(f)$.

Composition

- For any morphisms f and g with $\mathrm{cod}(f) = \mathrm{dom}(g)$, there is a composite morphism $g \circ f$.

Associativity

- For any morphisms f, g, h: $h \circ (g \circ f) = (h \circ g) \circ f$

Identity

- For each object A, there is an identity morphism id_A.
- For any morphism f: $\mathrm{id}_{\mathrm{dom}(f)} \circ f = f \quad \text{and} \quad f \circ \mathrm{id}_{\mathrm{cod}(f)} = f$

Finite Limits

Terminal Object

- There is an object 1 (terminal object) such that for any object A, there is a unique morphism $! : A \to 1$.

Pullbacks

- For any pair of morphisms $f : A \to C$ and $g : B \to C$, there exists a pullback square:

$$\begin{array}{ccc} P & \longrightarrow & B \\ \downarrow & & \downarrow{\scriptstyle g} \\ A & \xrightarrow{f} & C \end{array}$$

Topos-theoretic foundations

We include in our axioms for topos-theoretic foundations: NNO, Boolean, axiom of choice.

Topos-theoretic foundations

Element

For an object A in $\mathcal{E}$, an **element** of A is an arrow $x : 1 \to A$.

Intuitive differences between **Set** and **Cat**

In **Set**: any two sets can stand in the elementhood relationship with each other.

The question of framework

- Since both ZF and ETCS can be formulated in many-sorted, classical, first-order logic, they can be compared by standard tools (such as bi-interpretability).
- But there is a sense in which "thinking categorically" or "thinking type-theoretically" does not sit well within this framework.

Shulman's Theorem

Shulman (2019) seems very close to proving bi-interpretability of ZF and ETCS.

- For each model U of ZF, there is a corresponding model of ETCS; and for each model $\mathcal{E}$ of ETCS, there is a corresponding model of ZF.
- What are the permitted constructions?
- In what sense is the construction uniform, i.e. doesn't depend on specific features of a model?
- What needs to be shown about the constructions?

From universe to topos

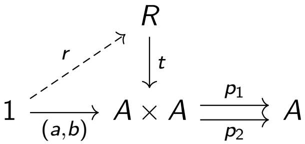
1. Given a model $\langle U, \in \rangle$ of ZF, let $\mathcal{E}_0 = U$, and let $\mathcal{E}_1$ be the set of functions between sets (constructed as subsets of ordered pairs).
2. Fact: the pair $\mathcal{E}_0, \mathcal{E}_1$ forms a model of ETCS.
    - The empty set is an initial object.
    - Any singleton set is a terminal object.
    - Etc.

From topos to universe

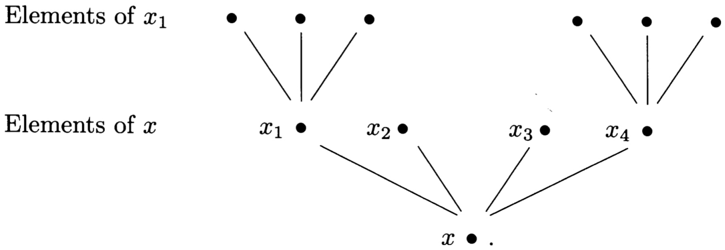
- Intuitively, the objects in $\mathcal{E}$ would become sets. But how to define the relation $A \in B$?
- So instead of taking objects in $\mathcal{E}$ as sets, we take trees:

$$t : R \rightarrowtail A \times A$$

For elements $a : 1 \to A$ and $b : 1 \to A$, we write $a \leq b$ just in case



Construction of ZF model from ETCS model



Tree: A **tree** is a poset that is downward linear.

Rooted: If $t : R \rightarrowtail A \times A$ is a tree, and $e : 1 \to A$, then we say that e is the **root** of t just in case $\forall x(e \leq x)$.

Accessible: A pointed tree (t, e) is **accessible** just in case: for every element $x : 1 \to A$ there is a finite R-path to the root $e : 1 \to A$.[4]

[4] This definition can be made first-order using subobjects of the natural number object in $\mathcal{E}$.

A subobject $m : S \rightarrowtail A$ is said to be **inductive** for the tree $t : R \rightarrowtail A \times A$ just in case: for any element $x : 1 \to A$, if every $y \leq x$ factors through m, then x factors through m.

Well-founded: If $m : S \rightarrowtail X$ is inductive, then m is an isomorphism.

Extensional: For any $x : 1 \to A$ and $y : 1 \to A$, if x and y have the same R-children, then $x = y$.

How do we know that there are “enough” of these trees in $\mathcal{E}$ to build an entire ZF universe?

Questions about Shulman's result

- The construction of trees from a topos $\mathcal{E}$ seems to require infinitary procedures. Is this move permitted by the standard definition of bi-interpretability?

A simple example

T_1 says that there are exactly two things.
T_2 says that there are exactly two atoms, and one mereological sum of those atoms.

To establish bi-interpretability, it needs to be shown that there is a formula χ (in the language of category theory) that defines an isomorphism between $GF(\mathcal{E})$ and $\mathcal{E}$.

Type theory: Kemeny or Awodey?

"It was my intention to prove the equivalence of the simple theory of types and Zermelo set-theory. Instead of this I have succeeded in proving a strong theorem from which it follows that the two systems are not equivalent under <u>any</u> reasonable definition of 'equivalent'." (Kemeny, 1949)

Questions and Conjectures

Is the following an example of mutually interpretable theories that are not bi-interpretable?

T_1 is the theory of a field with 2 elements.
T_2 is the theory of fields of characteristic 2.

It depends on what we mean by "bi-interpretable". If equality has to be translated strictly, then there is no translation from T_1 to T_2.

Questions and Conjectures

K. Williams argues that bi-interpretability is not strong enough:
`http://kamerynjw.net/2022/05/18/bi-interpretability.html`

Questions and Conjectures

- Dependent type theory is a more natural setting for theory of categories and the elementary theory of toposes.
  - Replace "isomorphism" with "equivalence".
- If ETCS and ZF are formalized in FOLDS (Makkai, 1995), does the equivalence result still hold?

Questions and Conjectures

- Does the category $\underline{\mathrm{hom}}(T, T')$ of translations from T to T' correspond to a certain category of functors between the syntactic categories C_T and $C_{T'}$?
- It's tempting to move to a purely categorical framework (e.g. replace theories with Boolean coherent categories) because of the nastiness dealing with variables, binding, and substitution. Would anything be lost by doing so? How would we translate results back to tell us something about theories?

References I

Andréka, Hajnal, Judit Madarasz, and István Németi (1994). "Mutual definability does not imply bi-interpretability". In: *Studia Logica* 53.3, pp. 353–378. DOI: 10.1002/malq.200410051.

Awodey, Steve (2009). "From sets to types to categories to sets". In: *Philosophical Explorations*. DOI: 10.1007/978-94-007-0431-2_5.

Barrett, Thomas William and Hans Halvorson (2016). "Morita equivalence". In: *The Review of Symbolic Logic* 9.3, pp. 556–582.

Feferman, Solomon (1969). "Set-theoretical foundations for category theory". In: *Reports of the Midwest Category Seminar III*. Vol. 106. Lecture Notes in Mathematics. with an appendix by G. Kreisel. Springer, pp. 201–247.

— (1977). "Categorical foundations and foundations of category theory". In: *Logic, Foundations of Mathematics, and Computability Theory*. Ed. by Robert E. Butts and Jaakko Hintikka. Dordrecht: Reidel, pp. 149–169.

References II

Freire, Alfredo Roque and Joel David Hamkins (2021). "Bi-interpretation in weak set theories". In: *The Journal of Symbolic Logic* 86.2, pp. 609–634.

Friedman, Harvey M and Albert Visser (2014). "When bi-interpretability implies synonymy". In: *Logic Group preprint series* 320, pp. 1–19.

Gajda, Adam, Michal Krynicki, and Lesław Szczerba (1987). "A note on syntactical and semantical functions". In: *Studia Logica* 46.2, pp. 177–185. DOI: 10.1007/bf00370379.

Halvorson, Hans (2019). *The Logic in Philosophy of Science*. Cambridge University Press. ISBN: 9781107527744. DOI: 10.1017/9781108596855.

Kemeny, John George (1949). "Type-theory vs. set-theory". PhD thesis. Princeton University.

References III

- Lewis, David K et al. (1986). *On the plurality of worlds*. Blackwell Oxford.
- Linnebo, Øystein and Agustín Rayo (2012). "Hierarchies Ontological and Ideological". In: *Mind* 121.482, pp. 269–308. DOI: 10.1093/mind/fzs050.
- Mac Lane, Saunders and Ieke Moerdijk (1992). *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Universitext. New York, NY: Springer-Verlag. ISBN: 978-0387977102. DOI: 10.1007/978-1-4612-0927-0.
- Makkai, Michael (1995). "First order logic with dependent sorts, with applications to category theory". In: *Preprint: http://www.math.mcgill.ca/makkai*, p. 136.
- Makkai, Michael and Gonzalo E Reyes (2006). *First order categorical logic: model-theoretical methods in the theory of topoi and related categories*. Springer. DOI: 10.1007/BFb0066201.

References IV

- Mathias, Adrian R. D. (2001). "The strength of Mac Lane set theory". In: *Annals of Pure and Applied Logic* 110.1-3, pp. 107–234. DOI: 10.1016/S0168-0072(00)00031-2.
- McLarty, Colin (2004). "Exploring categorical structuralism". In: *Philosophia Mathematica* 12.1, pp. 37–53.
- Mitchell, Barry (1965). *Theory of categories*. Academic Press.
- Pinter, Charles C. (1978). "Properties preserved under definitional equivalence and interpretations". In: *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 24.10, pp. 481–488. DOI: 10.1002/malq.19780241004.
- Shulman, Michael (2019). "Comparing material and structural set theories". In: *Annals of Pure and Applied Logic* 170.4, pp. 465–504. DOI: 10.1016/j.apal.2018.11.002.

References V

- Szczerba, Lesław (1977). "Interpretability of elementary theories". In: *Logic, Foundations of Mathematics, and Computability Theory: Part One of the Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science, London, Ontario, Canada-1975*. Springer, pp. 129–145.
- Van Benthem, Johan and David Pearce (1984). "A mathematical characterization of interpretation between theories". In: *Studia Logica* 43, pp. 295–303. DOI: 10.1007/BF02103292.
- Visser, Albert (2006). "Categories of theories and interpretations". In: *Logic in Tehran*. Vol. 26. Association for Symbolic Logic, pp. 284–341. DOI: 10.1017/CBO9780511712068.012.