Equivalence: A Program

Hans Halvorson

Oslo, June 2026

The same theory, presented differently

The **same content** under **different forms of presentation**

- date and place of presentation, font, language, notation, primitives

When human beings communicate, the content is wrapped up in various forms.

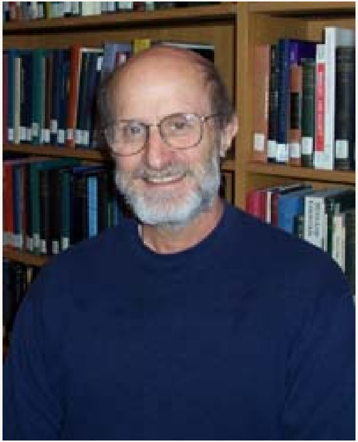
- Could one presentation say "there are exactly two things" while another presentation says "there are exactly three things"?
- Could one presentation (ZF) be based on a membership relation $x \in y$ while another presentation (ETCS) is based on morphisms $A \to B$ between sets?

Sixty years of technical work

Philosophical critics

> *Whatever this structural "isomorphism" is to be, it cannot be a purely formal notion. It cannot be, that is, an interrelationship which can be determined to hold solely on the basis of the logical form of the theories in question.*

— Sklar 1982



Philosophical critics

According to one alternative—the second "extreme" approach to equivalence I want to discuss—we can say that theories are equivalent without saying why they are equivalent in terms of fundamentality and underlying third languages. (Sider 2020)

I investigate whether there is a defensible view according to which formal criteria have significant non-mathematical implications, of these sorts or any other, reaching a chiefly negative verdict. (Teitel 2021)

Saving the program

1. Concession: No direct philosophical upshot of formal equivalence results.
2. Retrenchment: Equivalence results shed light on what combinations of philosophical commitments are tenable.
   - How much information transfers from one formulation to another
   - If you *insist* on one formulation, what does that say about your non-scientific philosophical commitments?

How logicians do it

The logician's criteria

Relations between first-order theories:

- mutual interpretability
- definitional equivalence
- bi-interpretability

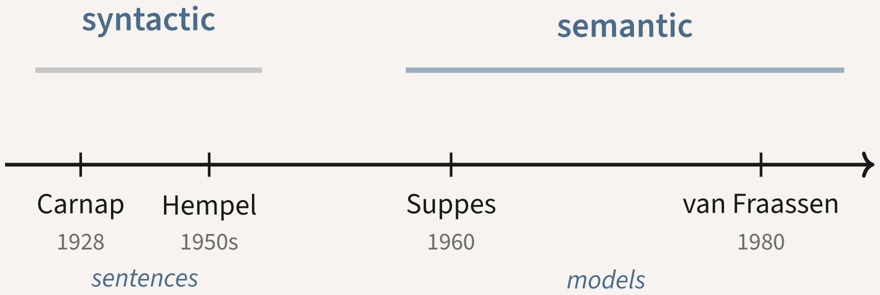
But physics is not written in first-order logic

These criteria presuppose a linguistic (first-order) regimentation.

The most intereresenting physical theories don't have canonical regimentations – general relativity uses Lorentzian manifolds, QFT uses operator algebras, Hamiltonian mechanics uses symplectic manifolds

The semantic view of scientific theories

In the 1960s and 70s, philosophers of science rejected the *syntactic* view of theories in favour of the *semantic* view.



When are two classes of models the same?

too Zeno

bijection of models

$\mathrm{Mod}(T)$ $\mathrm{Mod}(T')$

too Heraclitus

model-by-model isomorphism

$\mathrm{Mod}(T)$ $\mathrm{Mod}(T')$

The categorical move

A simple idea, naturally motivated by mathematical physics

- Philosophers of science could have found more plausible answers by keeping up with logic (see: relative interpretability)
- I found my way there by working backwards: the collection of models of a theory is **structured**:
    1. Morphisms
    2. Topology

A category

$$
\begin{tikzcd}
A \arrow[r, shift left, "f"] \arrow[dr, "h"'] & B \arrow[l, shift left, "g"] \\
& C
\end{tikzcd}
$$

The morphisms are the point: a category remembers how its objects relate.

A functor

$$
\begin{tikzcd}
A \arrow[r, dotted, "F"] \arrow[d, bend right, "f"'] \arrow[d, bend left, "g"] & X \arrow[d, bend right, "p"'] \arrow[d, bend left, "q"] \\
B \arrow[r, dotted] \arrow[d, "h"] & Y \\
C \arrow[ur, dotted, bend right] & Z \\
\mathbb{C} & \mathbb{D}
\end{tikzcd}
$$

Natural transformations

$$\begin{tikzcd}
& & & W \arrow[r, "Ff"] \arrow[dd, "\eta_A"'] & X \arrow[dd, "\eta_B"] \\
A \arrow[r] & B \arrow[urr, dashed, bend left, "F"] \arrow[drr, dashed, bend right, "G"'] \arrow[r, phantom, "\Downarrow\eta"] & & & \\
& & & Y \arrow[r, "Gf"'] & Z
\end{tikzcd}$$

Equivalent categories

Categories $\mathbf{C}$ and $\mathbf{D}$ are *equivalent* when there are functors F, G back and forth with $GF \cong 1_{\mathbf{C}}$ and $FG \cong 1_{\mathbf{D}}$.

$$\begin{tikzcd} A \arrow[d, "u"'] \\ B \end{tikzcd} \quad \underset{G}{\overset{F}{\rightleftarrows}} \quad \begin{tikzcd} W \arrow[r] \arrow[d, leftrightarrow, "\cong"'] & X \arrow[d, leftrightarrow, "\cong"] \\ Y \arrow[r] & Z \end{tikzcd}$$

$\mathbf{C}$ $\qquad\qquad\qquad\qquad$ $\mathbf{D}$

Now “same theory” has a precise, standard answer.

What is, and isn't, invariant

Preserved by categorical equivalence:

- categorical properties such as "has Cartesian products"
- relations between objects
- automorphism groups of objects

Not preserved:

- internal structures of objects
- number of objects

Categorical equivalence is too liberal

- Weatherall: physics depends on internal structure of models
- Hudetz: the equivalence should itself be definable
- A category doesn't **assert** anything

No downstairs without upstairs

To adopt a theory is to make assertions about objects — *downstairs*.

But it is also to commit, even if tacitly, to when a *different* formulation would be saying the same thing.

- that commitment is a criterion of equivalence — *upstairs*
- you cannot fix what you are asserting without it

There is no downstairs without an upstairs.

Different practices call for different criteria of equivalence

The physicist is rightly indifferent about foundational issues that the set-theorist is worried about.

The set theorist might be rightly indifferent about foundational issues that the philosopher of mathematics is worried about.

- not different *tastes* — different objects of inquiry
- each criterion is the exact expression of what aspect of reality its practice is trying to understand

No single criterion reads off the one true ontology.

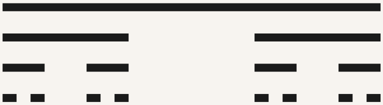
Sharpening the criterion

Categorical equivalence is too coarse

Distinct propositional theories can have equivalent (discrete) categories of models — yet differ in Stone space.

$\mathrm{Mod}(T)$

$p_0, p_1, \dots$ *no axioms*



Cantor space

$\mathrm{Mod}(T')$

$q_i \to q_j \; (i \leq j)$

Cantor space $\sqcup$ *isolated point*

Makkai's repair

Add structure to the category of models that the bare category forgets:

- **ultracategories** — categories equipped with ultraproduct functors
- Makkai's reconstruction: a theory can be recovered, up to equivalence, from its ultracategory of models
- the relation this induces: T and T' have equivalent *syntactic categories* / classifying toposes $\mathrm{Sh}(C_T) \simeq \mathrm{Sh}(C_{T'})$

Heavy apparatus — but it determines an interesting relation on first-order theories

The heavy thing is Morita equivalence

Barrett–Halvorson-Tsementzis: Makkai's relation is *exactly* this —

T and T' have a **common definitional extension**, allowing definitions of new **sorts** as well as new relation and function symbols.

That is **Morita equivalence**.

- the topos-theoretic apparatus and the syntactic notion coincide
- definitions of sorts are what categorical equivalence was silently using all along

Defining new sorts

Quotient

$$
\begin{tikzcd}
\sigma \arrow[d, "e"] \\
\sigma'
\end{tikzcd}
$$

Subsort

$$
\begin{tikzcd}
\sigma \\
\sigma' \arrow[u, "i"']
\end{tikzcd}
$$

Product

$$
\begin{tikzcd}
 & \sigma_0 \times \sigma_1 \arrow[dl] \arrow[dr] & \\
\sigma_0 & & \sigma_1
\end{tikzcd}
$$

Coproduct

$$
\begin{tikzcd}
 & \sigma_0 \amalg \sigma_1 & \\
\sigma_0 \arrow[ur] & & \sigma_1 \arrow[ul]
\end{tikzcd}
$$

Discovered twice

The same criterion was reached from the other direction – by logicians using no category theory at all.

- Szczerba 1977: Interpretability of elementary theories
  - Hodges *Model Theory*
  - Friedman and Visser
- the *default* notion of **bi-interpretability** coincides with Morita equivalence
- first proven by Evan Washington (Princeton undergraduate thesis, 2018); see *The Logic in Philosophy of Science*

Two approaches, one criterion.

A spectrum of criteria

How much do these results show?

Most equivalence results in philosophy of physics are proved at the categorical level:

- Hamiltonian and Lagrangian mechanics (Barrett 2019)
- general relativity and Einstein algebras (Rosenstock, Barrett & Weatherall 2015)

But if it's not sufficient for first-order theories, why assume it's sufficient for these theories?

Two levels

Accepting a theory involves two distinct acts:

1. Downstairs: Assertions we make about objects
2. Upstairs: Attitude toward our assertions
    - confidence levels
    - tolerance for reformulation

Quine-style naturalism has only downstairs. That's why it cannot understand what formal theories of equivalence are good for.

A program

- Formal results about equivalence are part of the classic tradition of critical philosophy – i.e. of reflecting on the rules that govern our scientific practice.
- Philosophers who don't recognize such rules are not going to find a relation of "equivalence" in the world.
- Kant said that scientific implies mathematical. That's why we want a "formal" account of equivalence; we want something that is intersubjectively accessible.
    - Make your theory explicit, and also make your criterion of equivalence explicit.