College *of* Arts *and* Letters
Notre Dame Philosophical Reviews

UNIVERSITY OF NOTRE DAME

2020.01.14

Everything

Hans Halvorson

The Logic in Philosophy of Science

Published: January 27, 2020

Hans Halvorson, *The Logic in Philosophy of Science*, Cambridge University Press, 2019, 296pp., $34.99 (pbk), ISBN 9781107527744.

Reviewed by Kevin Davey, University of Chicago

Many of the central issues of modern philosophy of science have explicitly or implicitly revolved around questions of when two theories should be viewed as 'equivalent' -- that is, when two theories are something like mere re-descriptions of one other. For example, the logical positivists took two theories to be equivalent whenever they make the same empirical predictions. By contrast, in the post-positivist tradition many philosophers have sought to draw distinctions between theories which are empirically equivalent, arguing that some are superior to others in virtue of the basic ontology they posit or other characteristics they possess. (We see this, for example, in debates about whether this or that interpretation of quantum mechanics or general relativity is to be preferred.) In this way, more fine-grained notions of theoretical equivalence have emerged, replacing older and more coarse-grained conceptions of theoretical equivalence. In a different debate, Putnam in his anti-realist argument has gone in the opposite direction and has allowed radical re-descriptions of the world to count as equivalent theoretical descriptions of reality, while Lewis and others have pushed back, in effect endorsing less generous notions of theoretical equivalence. These examples and others show the centrality of questions about equivalence of theories in the general philosophy of science.

As the reader will perhaps know, there is also a view about the nature of theories -- the so called *semantic view* of theories -- that resists thinking of theories as sets of sentences in a formal language, thinking of them instead as *models* or *sets of models*. But even on this view, one will still need to have some sort of story as to when two models (or sets of models) are in effect re-descriptions of each other. Thus, whether working from a more syntactic view of theories or a more semantic view of theories, broadly similar issues of equivalence of theory or identity of structure arise and quickly become central.

In spite of the central role that the idea of theoretical or structural equivalence has played in the philosophy of science, most philosophers have proceeded with only a crude, relatively informal understanding of this notion. The main goal of Hans Halvorson's book is to remedy this by presenting various formal accounts of theoretical equivalence, discussing their strengths and weaknesses. Perhaps most novel is the inclusion of category theory as a tool with which to have this discussion.

It is not hard to see why category theory belongs in such a discussion. Questions about the identity conditions for mathematical objects have undergone a paradigm shift with the development of category theory. Part of the motivation of category theory is the idea that to determine what makes a mathematical object what it is, we should not be trying to look inside the mathematical object at its intrinsic characteristics, but should rather focus on the relationship the mathematical object has with other mathematical objects. More specifically, we should look at the network of ways in which the mathematical object can be mapped into other similar mathematical objects, and that other mathematical objects can be mapped into it. One way of formalizing this is with the notion of the category of morphisms holding

between mathematical objects of a certain sort. Indeed, theories themselves can be viewed as mathematical objects that can be embedded into each other in various ways, as can models, and thus there arise various categories of theories, or categories of models. One might hope that broad notions of identity and equivalence of structure can be tackled by studying such categories of theories or models, and it is with this in mind that it is natural to use category theory as a tool in trying to articulate notions of theoretical equivalence. While Halvorson does not exclusively focus on category theoretic notions of theoretical equivalence, such notions nevertheless play a large role in this book, and their inclusion is perhaps one of the book's main novelties.

I now present a chapter by chapter breakdown of the contents of the book.

Chapter 1 is a review of the standard syntax and semantics of propositional logic, culminating in a definition of translatability of one theory into another and equivalence of theories (sometimes also known as homotopy equivalence) in the context of purely propositional theories.

Chapter 2 is an introduction to category theory, focusing on the development of the category of sets. While this chapter is technical, a good deal of it is not used in subsequent chapters. It might have been more useful to have presented a distilled discussion of only that part of category theory and the category of sets that is really relevant for what follows, referring the reader elsewhere for details. Still, the discussion is both useful and thorough.

Chapter 3 synthesizes the material of the previous two chapters by exploring the category of propositional theories whose arrows are translations between theories (as defined in chapter 1). This culminates in the result that the category of propositional theories thus defined and the category of Boolean Algebras are equivalent, and in turn dual to the category of Stone spaces. It is also proven that any consistent, finitely axiomatizable purely propositional theory in a countable signature is homotopy equivalent to the empty theory, creating connections with issues discussed by Quine.

Chapter 4 is a review of the standard syntax and semantics of first order logic. This leads to a definition of strong intertranslatability of theories (also called homotopy equivalence), formalizing the intuitive relation that holds between two theories when the formulae of one can be translated into the other, and vice versa. This generalizes the earlier definition of translatability given in the purely propositional context. A separate notion of definitional equivalence is also introduced, according to which two theories are equivalent when they have a common definitional extension. It is proven that definitionally equivalent theories are strongly intertranslatable. The converse result, that strongly intertranslatable theories are definitionally equivalent, is proved in chapter 6. That these intuitive conceptions of theoretical equivalence (strong intertranslatability and definitional equivalence) are provably equivalent shows that they do indeed correspond to a natural conception of theoretical equivalence.

While strong intertranslatability gives us a natural way to think of theoretical equivalence, this notion has its limitations. In particular, once one allows many-sorted theories, this notion of theoretical equivalence does not allow us to view as equivalent theories which intuitively seem equivalent. There is, however, a straightforward way of generalizing this notion to the many-sorted context. The resulting notion of theoretical equivalence is known as Morita equivalence, and this is one of the main subjects of chapter 5. It is proven in this chapter that any many-sorted theory is Morita equivalent to some single-sorted theory, which was of course the basis of Quine's critique of many-sorted logic. A separate notion of weak intertranslatability of theories (also a little confusingly given the same label of homotopy equivalence) is introduced, which turns out to be equivalent to Morita equivalence of theories (though this result is not proved until chapter 7.) That these intuitive conceptions of theoretical equivalence (weak intertranslatability and Morita equivalence) are provably equivalent shows that they too correspond to a natural conception of theoretical equivalence. Strongly intertranslatable (i.e., definitionally equivalent) theories are weakly intertranslatable (i.e., Morita equivalent), though not necessarily vice versa. So at this point there are two natural notions of theoretical equivalence on the table.

In chapter 6, the subject changes to models and semantics. After presenting some standard results of first-order model theory, the discussion turns to various categories of models. This leads to yet another natural notion of equivalence of theories, so-called categorical equivalence, which holds of two theories when their categories of models (with elementary embeddings as arrows) are equivalent. It is shown that weakly intertranslatable theories are categorically equivalent. Thus, categorical equivalence is a coarser relation than weak or strong intertranslatability. The chapter also includes a discussion of Beth's theorem and Svenonius' theorem, both of which revolve around semantic notions.

In chapter 7, the notion of model is generalized to the many-sorted case, along with the corresponding notion of equivalence of categories of models (i.e., categorical equivalence). After returning to the concept of Morita equivalence and exploring it in some detail, it is shown (among other things) that Morita equivalence entails categorical equivalence, but that categorical equivalence does not entail Morita equivalence. This rounds out the discussion of the relationship between the three main notions of theoretical equivalence on the table. The discussion then turns to geometry, and it is shown that various point-based theories of Euclidean, projective and affine geometry are Morita equivalent to various line-based theories of these geometries.

Chapter 8 involves an application of the tools and concepts developed in the previous chapters to a disparate set of purely philosophical issues. Because of its philosophical interest, I describe it in considerably more detail. The chapter begins with a discussion of Ramsey sentences and their role in the philosophy of science. One might try to formulate a notion of theoretical equivalence according to which two theories are equivalent if and only if their Ramsey sentences are logically equivalent (according to some conception or other of second-order logical equivalence.) Problems are raised for this notion of theoretical equivalence, lending support to the old idea that Ramsey sentences are 'too easily' made true. This is all intended to count as an argument against Ramsey sentence structuralism.

The discussion then moves to we should say about the relationship between isomorphic models. Halvorson argues that in postulating a theory T, we should not view ourselves as taking sides as to whether two given isomorphic models of T are identical. More abstractly, the view advanced is that commitment to T does not require us to say anything not invariant under categorical equivalence of the category of models of T (with elementary embeddings as arrows.) Any claims beyond this -- such as claims about whether certain models should be understood as distinct or not -- need to be understood as meta-theoretical claims that go beyond our commitment to T.

Halvorson then goes on to critique Putnam's famous model-theoretic argument against realism. He presents Putnam's argument as one concerning theoretical equivalence, according to which (under minimal assumptions) all consistent theories are equivalent. Halvorson's criticism of Putnam's argument covers much ground, but I will summarize what I take to be its core claim.

Putnam's argument presupposes that our theory, T, of the world has a fixed interpretation that can be spelled out in set-theoretic terms in the usual way (i.e., with a domain, a subset of the domain for each unary predicate, and so on). Putnam's argument then requires that he be able to make semantic ('meta-theoretic') claims about the relationship between T and this interpretation. For example, Putnam's argument presupposes that he can make claims like 'this speaker uses *t* to refer to *o*', where *t* is a term in the language of T, and *o* is an object in the interpretation. Such claims, however, are not a consequence of our commitment to T (and indeed, are not even expressible in the language of T).

Putnam's argument therefore takes for granted that there is a kind of meta-theoretic point of view from which we can make certain factual claims about which terms in our language denote which objects in our interpretation, and so on. For Putnam's argument to work, claims in this meta-language must have determinate truth values. That is to say, it cannot be the case that claims like 'this speaker uses *t* to refer to *o*' and 'it is false that this speaker uses *t* to refer to *o*' have an equal claim to truth, otherwise Putnam's argument cannot proceed. Thus, Putnam must assume that his own argument fails when it comes to meta-level discourse. He must assume that the meta-theorist's language 'gets a grip on the world' in a way in which that of the object-level theorist does not. But of course, metatheory is really just more theory -- and thus

Putnam's argument undermines itself. Halvorson makes this point several ways, summarizing it as the claim that Putnam (problematically) assumes that the world can be described as an object in the category Sets. (It would have been satisfying to have seen a clearer statement of the connection, if one exists, between all this and the idea stated immediately prior that commitment to T does not require us to say anything not invariant under categorical equivalence of the category of models of T.) Halvorson then distinguishes his criticism of Putnam from that of Lewis. All in all, this leaves us with an interesting reply to Putnam's argument that can be added to the many others already in print.

In addition to this argument, Halvorson points out several times that all Putnam's argument shows us is that a rival theory T' *could* be charitably interpreted as true -- but that this does not mean that we *must* judge T' as literally true. (This is surely a distinct objection from that just given.)

Halvorson goes on to discuss realism, arguing that what distinguishes the realist from the antirealist is that the antirealist is willing to call many theories equivalent that the realist is not. Thus, the dispute between the realist and antirealist is one about the equivalence of theories, underscoring the centrality of this concept in the philosophy of science. (This general point about realism is made several times in the book.) Given this, one might hope that the 'correct' notion of theoretical equivalence could then be used to settle all sorts of metaphysical disputes. However, Halvorson resists the idea that we must make a choice between the various notions of theoretical equivalence developed over the course of the book -- his overall view is a more pragmatic one according to which each notion of theoretical equivalence is a tool that is useful in some contexts, and not in others, perhaps depending on our antecedent attitudes to the example under discussion.

Halvorson concludes by arguing that scientists are not typically interested in representing the world with single models, but with structured sets of models that are connected with each other in various ways (e.g., in a category or a topology). He thus concludes that our attitude to scientific theories is quite different than mere belief in some set of propositions that we take to be true, or mere belief that some model is isomorphic to reality. In this sense, simple versions of the semantic theory (one of the book's main foils) cannot be right -- though surely no simplistic version of the syntactic theory can be right either.

Having summarized the book's contents, I make a few general remarks on its materials and approach in no particular order.

The reader should be aware that Halvorson's goal in this book is not to present a single sustained view on the nature of theoretical equivalence (even though distinct points of view do emerge), but rather to introduce and develop various formal tools with which discussions of theoretical equivalence can proceed more rigorously.

The book is very technical, and covers much ground. Although it reviews material of a more elementary nature (such as the soundness and completeness of first-order logic), much of the material is developed from a more abstract point of view, and so is probably not suitable for readers without a fairly high level of mathematical sophistication. On a critical note, distinctions are not always drawn between technical material on which subsequent arguments depend and technical material which is not developed further (for example, very little of the detail on the category of sets is subsequently used). This can make the book's main thrust of hard to discern at times, particularly given that certain important equivalences -- e.g., between strongly intertranslatable and definitionally equivalent theories, and weakly intertranslatable and Morita equivalent theories -- are proved over the course of several chapters. The terminology can also be slightly confusing, as for example when the term homotopy equivalence is used to describe several distinct notions. So on a technical level, it is a book which places significant demands on the reader. But on a positive note, those who can patiently work through it will find themselves greatly rewarded by the end with an elegant mathematical picture of various notions of theoretical equivalence and their relations to one another.

No doubt readers will be divided on the question of whether the philosophical payoff for all the technical work of the earlier chapters is sufficient. While of great interest, much of the purely philosophical material of chapter 8 actually requires very few of the technicalities developed in the previous chapters. The same is true to some extent of the earlier discussion of implicit and explicit definability and its connection with the philosophy of mind. This is not intended as a criticism, but is rather just to raise the question of what is at stake in the book's main issues when things are really boiled down to their core. Hopefully future work by Halvorson and others will make this clearer.

One point that is especially intriguing is Halvorson's idea that categorical equivalence -- the notion of theoretical equivalence developed in chapter 6 -- is motivated '*not by linguistic considerations, but by scientific practice*', and by focusing on what scientists '*actually do*' with theories. This point is only developed and defended in cursory detail, and the question of how persuasive a case can be made for it is of great interest. Hopefully future work can shed more light on this. A natural question is: if categorical equivalence really is the notion of theoretical practice most closely motivated by scientific practice, why should it not be viewed as the 'correct' notion of theoretical equivalence, and other notions of theoretical equivalence abandoned? More could be said about this, but I will not pursue this point further here.

It is also worth noting -- as Halvorson does on p. 219ff -- that the kind of toy examples of first-order scientific theories that fill the book are very different from the more complex theories found in actual science, many of which do not admit a first-order formalization. (Halvorson notes that neither the general theory of relativity or quantum mechanics are first order formalizable.) Thus, there remains a great distance between the questions about theoretical equivalence that can be resolved by the sorts of techniques developed in the book, and many of the questions about theoretical equivalence that one finds in modern philosophy of science. We can only hope that further developments will narrow this gap, but for now, the gap remains significant.

None of these cautionary remarks are intended to detract from Halvorson's wonderful accomplishment. The book is large in scope, and can be said to begin in earnest a much-needed rigorous study of the concept of equivalence of theories. There is much more on the topic that needs to be said, but without a seminal text of this sort such conversations can struggle to even begin. Anyone interested in general issues in metaphysics and the philosophy of science will find much to be stimulated by here.

ISSN: 1538 - 1617