

Categories of scientific theories

Hans Halvorson and Dimitris Tsementzis*

July 29, 2015

Abstract: We discuss ways in which category theory might be useful in philosophy of science, in particular for articulating the structure of scientific theories. We argue, moreover, that a categorical approach transcends the syntax-semantics dichotomy in 20th century analytic philosophy of science.

Keywords: categorical logic, topos theory, semantic view of theories

Our aim in this article is to recommend category theory to philosophers of science, in particular as a means to articulating the structure of scientific theories. We are not suggesting that we replace first-order logic, model theory, set theory, and similar formal tools with category theory — as if category theory were just one more competitor among various formal approaches to philosophy of science. Much less are we proposing to replace set theory with category theory as the foundation of mathematics. Rather, we suggest that category theory unifies various approaches to formal philosophy of science, and shows that some of the debates between various approaches have been misguided. But most importantly, our proposal is not ideological, i.e. we have no stake in the claim that category theory is the “one and only correct” approach to scientific theories, much less that, “a scientific theory *is* a category.” Rather, we are merely sketching a program of research in formal philosophy of science: we suggest that it might be interesting to think of the “universe” of scientific theories as a *category of categories*, or more precisely, as a *2-category of categories*.

Our proposal includes the idea that familiar scientific theories (e.g. Hamiltonian mechanics, special and general relativity, quantum mechanics, quantum field theory) can themselves fruitfully

*Department of Philosophy, Princeton University

be described as categories.¹ If we represent theories this way, then we can take philosophical questions — e.g. are Hamiltonian and Lagrangian mechanics equivalent theories? — make them precise, and then use mathematical tools to answer these questions. We can also suggest various explications of important notions, such as *equivalence* or *reducibility*, and can then try to prove general theorems about such notions. In short, while we make no argument that we ought to use category theory, we would like to convince philosophers that category theory opens up a treasure trove of technical projects.

1 Theories as categories

Before we begin to discuss scientific theories, we review some basic notions of categorical logic, in order to frame the discussion of theories as categories. There are two salient ways in which a theory can be thought of as a category — a syntactic way, and a semantic way.

1.1 The syntactic category

In this entire chapter, when we speak of first-order logic, we mean first-order logic with possibly many sorts. Allowing the flexibility of many sorts doesn't truly add to the expressive power of first-order logic; but ignoring the possibility of many sorts can lead to needless confusions (see Barrett and Halvorson, 2015c).

What we mean by “many sorted” logic is that a signature Σ comes with a (finite) list $\sigma_1, \sigma_2, \dots$ of types, and variables, quantifiers, etc. are tagged by a particular type. For example, for each type σ , there is an equality symbol $=_\sigma$, which can be applied only to terms of type σ . Similarly, each predicate symbol $p \in \Sigma$ has an **arity** $\sigma_1 \times \dots \times \sigma_n$, where $\sigma_1, \dots, \sigma_n \in \Sigma$ are (not necessarily distinct) sort symbols. Likewise, each function symbol $f \in \Sigma$ has an **arity** $\sigma_1 \times \dots \times \sigma_n \rightarrow \sigma$, where $\sigma_1, \dots, \sigma_n, \sigma \in \Sigma$ are again (not necessarily distinct) sort symbols. Lastly, each constant symbol $c \in \Sigma$ is assigned a sort $\sigma \in \Sigma$. In addition to the elements of Σ we also have a stock of variables. We use the letters x, y , and z to denote these variables, adding subscripts when necessary. Each variable has a sort $\sigma \in \Sigma$.

Given a signature Σ , we define the terms and formulas of Σ in the normal way (see Barrett and Halvorson, 2015b). A *theory* T in Σ , in the sense of first-order logic is a set of sentences (or

¹For more on this idea, including evidence of its fruitfulness, see (Weatherall, 2016).

sequents) of Σ . There are, of course, many well known examples of such theories: e.g. the theory of partially ordered sets, the theory of groups, the theory of Boolean algebras, the theory of vector spaces over a field, the theory of categories, etc..

The immediate goal is to associate a category C_T with a theory T . For several reasons, we will suppose that the theory T is formulated in the coherent fragment of first-order logic, whose only connectives are \wedge and \vee , and whose only quantifier is \exists .^{2,3} The standard way of building a “syntactic category” for T is described in many works on categorical logic — see (MacLane and Moerdijk, 2012, p. 555), (Makkai and Reyes, 1977, p. 241), (Johnstone, 2002, p. 841), and (van Oosten, 2002, p. 39). In outline: an object of the syntactic category C_T is a formula in context, i.e. if ϕ is a formula of Σ , and if \vec{x} is a string of variables containing all those free in ϕ , then $\{\vec{x}.\phi\}$ is a formula in context. (Note that the objects of C_T depend only on the signature Σ , and not on the theory T .) Defining the arrows for C_T takes a bit more work. Let $\chi(\vec{x}, \vec{y})$ be a formula of Σ , where \vec{x} and \vec{y} are mutually disjoint sequences of variables. We say that $\chi(\vec{x}, \vec{y})$ is a T -provably functional relationship from $\{\vec{x}.\phi\}$ to $\{\vec{y}.\psi\}$ just in case T entails that, “for any \vec{x} such that $\phi(\vec{x})$, there is a unique \vec{y} such that $\psi(\vec{y})$ and $\chi(\vec{x}, \vec{y})$.” (The precise definition can be found in the aforementioned works on categorical logic.) Then we define an arrow from $\{\vec{x}.\phi\}$ to $\{\vec{y}.\psi\}$ to be an equivalence class, relative to T provable equivalence, of T -provably functional relations from $\{\vec{x}.\phi\}$ to $\{\vec{y}.\psi\}$.

The idea of a syntactic category might seem abstract and unfamiliar. But it’s a direct generalization of the more familiar idea of a Lindenbaum algebra from propositional logic. Suppose that $\Sigma = \{p_0, p_1, \dots\}$ is propositional signature, and let T be a theory in Σ . In this case, a formula in context simplifies to a sentence; and so the objects of C_T are just sentences. In this case, there is one provably functional relation (up to T -provable equivalence) between ϕ and ψ just in case $T, \phi \vdash \psi$ and otherwise there is no such provably functional relation. In other words, in C_T for a propositional theory T , there is an arrow from ϕ to ψ just in case $T, \phi \vdash \psi$.

Thus, from a theory T (considered as a set of sentences in Σ) we have constructed a category C_T .

²Strictly speaking this means that a theory T is a set of sequents of the form $\phi \vdash \psi$ where ϕ and ψ are coherent formulas. Alternatively, one can understand such sequents as first-order sentences of the form $\forall \vec{x}(\phi \rightarrow \psi)$ (where \vec{x} includes the unbound variables of both ϕ and ψ). In order not to deviate too much from the standard notation familiar to philosophers and logicians alike we will consider coherent theories to be sets of first-order sentences of the above-described form – this will allow us to speak of truth and satisfaction of a *sentence* rather than of a *sequent*, which is closer to the standard way of thinking about these matters.

³There is reason to think that the coherent fragment is adequate to formulate any theory that can be formulated in full first order logic. In particular, via Morleyization, every first-order theory is Morita equivalent to a coherent theory (see Tsementzis, 2015). What’s more, we agree that coherent logic is special: “there are good reasons why it is better to take $L_{\omega\omega}^c$ as basic rather than $L_{\omega\omega}$.” (Makkai and Reyes, 1977, p. 121).

And note that C_T partially eliminates the “language dependence” of T , which was so bemoaned by advocates of the semantic view of theories. Indeed, while T is bound to a particular signature Σ , the syntactic category C_T is independent of signature in the following sense: two theories T and T' , in different signatures, can nonetheless have equivalent syntactic categories.

But is the syntactic category C_T an adequate representative of the original theory T ? Here we answer in the affirmative, following Makkai and Reyes:

“In Chapter 8 we will show that, in a sense made precise there, logical [i.e. coherent] categories are *the same* as theories in a finitary coherent logic $L_{\omega\omega}^g$.” (Makkai and Reyes, 1977, p. 121)

And also:

“The content of 8.1.3 and 8.1.4 can be expressed by saying that for all practical purposes, T and C_T are the same.” (Makkai and Reyes, 1977, p. 241)

In what sense are T and C_T the same? There are a couple of ways we could answer this question. First, the theory T can be reconstructed from C_T in the following sense: each coherent category C gives rise to a canonically specified (coherent) theory T_C . T_C is (essentially) the set of sentences satisfied by C when C is understood as a model of the theory of coherent categories, i.e. the “total” theory of C qua coherent category. More precisely, for any coherent category C we have its *canonical language* Σ_C whose sorts are the objects of C and function symbols the arrows of C (sorted in the obvious way). Over this language Σ_C we can then express in a straightforward way what it is for a diagram in C to commute, what it is for a diagram to be a product diagram etc. T_C is then the collection of all those Σ_C -sentences expressing all those facts that are true of C as a coherent category. We then have:

Theorem 1. *Given a theory T , and its syntactic category C_T , the internal theory T_{C_T} of C_T is Morita equivalent to T .*

For the proof, see (Tsementzis, 2015, Corollary 4.6). We will further explain the notion of Morita equivalence in the following subsection as well as argue for its suitability as a good notion of equivalence between theories. For now, one may simply read the above result as “ T is equivalent to T_{C_T} ” and take it on faith that “Morita equivalence” is a sensible notion of equivalence between theories.

Second, the category $\text{Mod}(T)$ of models of T can be reconstructed from its syntactic category C_T .

Theorem 2. *Let T be a coherent theory, and let C_T be its syntactic category. Let $\mathbf{Coh}(C_T, \mathcal{S})$ be the category whose objects are coherent functors from C_T into the category \mathcal{S} of sets, and whose arrows are natural transformations. Let $\text{Mod}(T)$ be the category whose objects are models of T , and whose arrows are homomorphisms between models. Then $\mathbf{Coh}(C_T, \mathcal{S})$ is equivalent to $\text{Mod}(T)$.*

For the proof, see (Makkai and Reyes, 1977, p. 240).

The two preceding results show the sense in which there is no loss of essential information in passing from a theory T to its syntactic category C_T . (In particular, the Morita equivalence class of a theory can be recovered from that theory’s syntactic category.) It is tempting now to conjecture that if T and T' are Morita equivalent, then C_T and $C_{T'}$ are equivalent categories. But that conjecture fails, since the categories C_T and $C_{T'}$ might not be “conceptually complete” in the sense of Makkai and Reyes (1977). We discuss this issue further in the following subsection.

1.2 Equivalent theories

Before proceeding to discuss the semantic category $\text{Mod}(T)$ associated with a theory T , we will briefly discuss some ideas about when two theories are *equivalent* (for further discussion and technical results, see (Barrett and Halvorson, 2015a,b; Tsementzis, 2015)). The first question to be asked here is what notion of equivalence are we intending to capture? Our answer here is that we have no intention of capturing any Platonic essence of “equivalence.” Rather, just as a group theory gives us a fruitful notion of equivalence between groups (viz. isomorphism), and just as category theory gives us a fruitful notion of equivalence between categories (viz. categorical equivalence), so when theories are treated as mathematical objects, we hope to find a notion of equivalence that will be useful and illuminating.

The strictest notion of equivalence between theories is logical equivalence: two theories T and T' are said to be *logically equivalent* just in case they are formulated in the same signature Σ , and they have the same logical consequences among the sentences of Σ . Of course, logical equivalence is of no use for theories formulated in different signatures. For that case, we look to notions of how a theory can define new concepts that do not occur in the original signature Σ . Recall that a definitional extension T^+ in Σ^+ of the theory T in Σ is the result of adding new predicate symbols,

function symbols, or constant symbols that can be defined by T in terms of formulas in the original signature Σ . There is every reason, moreover, to think of a definitional extension T^+ as equivalent to the original theory T . Thus, two theories T_1 (in Σ_1) and T_2 (in Σ_2) are said to be *definitionally equivalent* just in case there are definitional extensions T_i^+ of T_i in $\Sigma_1 \cup \Sigma_2$ (for $i = 1, 2$) such that T_1^+ is logically equivalent to T_2^+ .

Nonetheless, there are reasons to think that definitional equivalence is not the most fruitful notion of equivalence between theories. One such reason is that it doesn't match well with the notions of equivalence between the corresponding syntactic categories (two theories can have equivalent syntactic categories without being definitionally equivalent). Another reason is that definitional equivalence cannot capture the sense in which, for example, the theory of categories can be equivalently formulated using objects and arrows, or just with arrows. In order to capture these intuitive verdicts of equivalence, the most plausible idea is *Morita equivalence*, which allows for equivalence of theories formulated not only in different signatures but also in different signatures with different sorts.

The notion of “Morita equivalence” of theories has two independent sources.⁴ On the one hand, Morita equivalence is suggested by ideas from categorical logic, in particular from topos theory. To see this, recall that the *pretopos completion* $P(C)$ of a coherent category C can be described in the following equivalent ways:

1. $P(C)$ is the result of freely adjoining finite coproducts and coequalizers of equivalence relations to C (see Johnstone, 2002, A1.4).
2. $P(C)$ is the subcategory of coherent objects in the topos $\text{Sh}(C)$ of sheaves on C , where the site C is equipped with the coherent Grothendieck topology.

Recall also that the *classifying topos* \mathcal{E}_T of the theory T is the unique (up to categorical equivalence) topos that contains a model of T , and such that any model of T in another topos \mathcal{E} uniquely lifts to a geometric morphism from \mathcal{E}_T into \mathcal{E} (see MacLane and Moerdijk, 2012, p. 561 and Makkai and Reyes, 1977, p. 272).

Now two coherent theories S and T are said to be *Morita equivalent* (in the categorical sense) just in case the following equivalent conditions hold (see Johnstone, 2002).

⁴The name “Morita equivalence” originates in module theory, and was transmitted into category theory through the study of algebraic theories.

1. The classifying toposes \mathcal{E}_S and \mathcal{E}_T are equivalent.
2. The pretoposes P_S and P_T are equivalent.⁵

The equivalence between these two statements follows from the fact that $\mathcal{E}_T \simeq \text{Sh}(C_T) \simeq \text{Sh}(P_T)$. Moreover, it follows from the second fact that if C_T is equivalent to C_S , then S and T are Morita equivalent. To see that the converse is not true, it suffices to display a theory T such that its syntactic category C_T is not a pretopos. Such theories are easy to find (see Example 1 below).

In a completely unrelated development, ideas related to Morita equivalence began to spring up in the works of logicians. As noted by Harnik (2011), Shelah’s T^{eq} construction (also known as “elimination of imaginaries”) is closely related to the pretopos completion construction. Indeed, the pretopos completion P_T of a theory C_T is the same thing as the syntactic category of T^{eq} , i.e. Shelah’s construction applied to T . Similarly, Andr eka et al. (2001) generalize the notion of a definitional extension so as to include the possibility of defining new sort symbols. We refer the reader to those works to see the original motivations for moving to a more expansive notion of equivalence between theories.

But what exactly does this all mean at the level of syntax? As has been explained by Barrett and Halvorson (2015b) and Tsementzis (2015), the ideas about Morita equivalence coming from topos theory correspond to a completely natural generalization of the idea of having a common definitional extension. In particular, given a theory T in signature Σ , a *Morita extension* T^+ of T can be constructed either by defining new relation and/or function symbols, or by defining new sorts from the sorts of Σ . The operation of defining new sorts via T corresponds roughly to taking the pretopos completion of C_T . Thus, intuitively speaking, two theories T and T' are Morita equivalent just in case T can define all the sorts, relation symbols, etc. of T' , and vice versa, in a compatible fashion.⁶ And in fact it can be shown that two theories are Morita equivalent in the syntactic sense just in case the pretopos completions of their syntactic categories are equivalent – see (Tsementzis, 2015, Theorem 4.7). This merely expresses the fact that the syntactic notion of Morita equivalence developed in (Barrett and Halvorson, 2015b) coincides with – and therefore characterizes – the topos-theoretic notion defined in (Johnstone, 2002). This justifies our free use of the same term “Morita equivalence” to refer to both notions.

⁵Here we use P_T to abbreviate $P(C_T)$, the pretopos completion of the syntactic category of T .

⁶For an example of how this “definitional” understanding of Morita equivalence can be applied to issues of theoretical equivalence in physics, see the discussion on classical mechanics in (Teh and Tsementzis, 2015).

Clearly if two syntactic categories C_T and $C_{T'}$ are equivalent, then T and T' are Morita equivalent. The converse, however, is not true.

Example 1. An easy way to see this is to take the an empty two-sorted theory T , i.e. the theory with no axioms whose signature Σ consists only of two sort symbols σ_1, σ_2 . Then we can extend T to T' by adding a “coproduct sort” $\sigma_1 + \sigma_2$ together with function symbols $\rho_1: \sigma_1 \rightarrow \sigma_1 + \sigma_2$ and $\rho_2: \sigma_2 \rightarrow \sigma_1 + \sigma_2$ and axioms defining $\sigma_1 + \sigma_2$ as a “coproduct” with ρ_1 and ρ_2 as its coprojections. Indeed T' is exactly a Morita extension of T in the sense of (Barrett and Halvorson, 2015b) which means that T and T' are Morita equivalent. However, the syntactic categories C_T and $C_{T'}$ cannot be equivalent: the obvious embedding $C_T \hookrightarrow C_{T'}$ is full and faithful but there can be no isomorphism from $\{z: \sigma_1 + \sigma_2. \top\}$ to any object of C_T (regarded as a full subcategory of $C_{T'}$). (Given the results of Tsementzis (2015), this is also an example of two theories which have equivalent pretopos completions but inequivalent syntactic categories.) \square

As such, we are left with two distinct notions of equivalence between theories T and T' :

(SE) Equivalence of their syntactic categories, i.e. $C_T \simeq C_{T'}$

(ME) Morita Equivalence

For reasons too detailed to go into here, we believe that (SE), although weaker than logical equivalence, is still too strong a notion. As noted above, (SE) is a sufficient condition for (ME) to hold, but not a necessary one. There is a strong sense in which (ME) captures exactly the right content of a theory as long as we care about that theory only up to the structure of its category of models. In order to clarify this remark, we must now go on to say a few more words about this category of models.

1.3 The semantic category

We have already noted that there is a second category associated with a theory T , namely the category $\text{Mod}(T)$ of its models. We will call $\text{Mod}(T)$ the *semantic category* associated with T .

Before proceeding, let’s be more precise about what we mean by the category of models of T . The objects of this category are simply set-valued models of T , in the sense of Tarski. But what are the arrows of the semantic category? There are two possible choices:

- Let T be a theory in signature Σ . We let $\text{Mod}(T)$ denote the category whose objects are Σ -structures that satisfy T , and whose arrows are homomorphisms of Σ -structures. Recall that if M and N are Σ -structures, then a homomorphism $j : M \rightarrow N$ is a function that preserves the extensions of symbols in Σ . That is, for each relation symbol $r \in \Sigma$,

$$j(r^M) \subseteq r^N, \tag{1}$$

and so on.

In contrast to the definition found in most model theory textbooks, we do *not* require the map j to be one-to-one, nor do we require equality in (1). The reason we don't impose these requirements is because they are unmotivated when the logic at hand doesn't have a negation symbol (as in the case of coherent logic).

For example, if $\Sigma = \{\circ, e\}$, and if T is the theory of groups (written in Σ), then the notion of a homomorphism of Σ -structures is simply the notion of a group homomorphism.

- We let $\text{Mod}_e(T)$ denote the category whose objects are (again) Σ -structures that satisfy T , and whose arrows are elementary embeddings of Σ -structures. Recall that if M and N are Σ -structures, then an elementary embedding $j : M \rightarrow N$ is a function that preserves extensions of all Σ -formulas. That is, for any formula $\phi(\vec{x})$ of Σ , and for any n -tuple \vec{a} of elements of M ,

$$M \models \phi(\vec{a}) \implies N \models \phi(j(\vec{a})). \tag{2}$$

In particular, for any sentence ϕ of Σ ,

$$M \models \phi \implies N \models \phi. \tag{3}$$

We can think of $\text{Mod}(T)$ as the “thick” category of models (more arrows) and $\text{Mod}_e(T)$ as the “thin” category of models (fewer arrows) of T . Note that $\text{Mod}_e(T)$ is a subcategory of $\text{Mod}(T)$, and typically a proper subcategory.⁷ For example, let Σ be a signature with one sort and no non-logical vocabulary, and let T be the empty theory in Σ , i.e. the theory whose models are bare sets. Let m_i

⁷Indeed, $\text{Mod}_e(T) = \text{Mod}(T)$ if and only if every first-order formula ϕ (over the signature Σ of T) is T -provably equivalent to a coherent formula (over classical logic) – see Johnstone (2002, Proposition 3.4.9).

be a model of T with i elements. Then $\text{Mod}(T)$ has an arrow $j : m_1 \rightarrow m_2$, whereas $\text{Mod}_e(T)$ has no such arrow (since elementary embeddings preserve the truth-value of numerical statements).

There are several questions one can ask of the relation between thin and thick categories. A well-known fact – alluded to above – is that the thin category of a full first-order theory is always equivalent to the thick category of a coherent theory, called its *Morleyization* – see (Johnstone, 2002, Lemma D1.5.13). There are also many natural open questions ripe for investigation regarding the relationship between thick and thin categories. We will mention some of them in the following section.

But for now, the most important point of this subsection is that $\text{Mod}(T)$, and *a fortiori* $\text{Mod}_e(T)$, is *not* generally an adequate representative of the theory T . This means that even the thick categories of models are not “thick enough”. To be more precise, the passage from T to $\text{Mod}(T)$ loses information in the sense that neither T , nor a theory T' that is Morita equivalent to T , can be reconstructed from $\text{Mod}(T)$. Examples from propositional logic makes this fact clear.

Example 2. For full first-order theories, there is an intuitive example. Let Σ be the propositional logic signature with symbols p_0, p_1, \dots . Let T_1 be the empty theory in Σ , and let T_2 be the theory with axioms $p_0 \vdash p_i$ for all $i \in \mathbb{N}$. Clearly T_1 and T_2 are *not* Morita equivalent theories. And yet, the semantic categories $\text{Mod}(T_1)$ and $\text{Mod}(T_2)$ are equivalent — since both are discrete, and have 2^{\aleph_0} objects. Thus, T_1 and T_2 are inequivalent theories whose semantic categories are equivalent (see Halvorson, 2012, p. 191). ┘

Example 3. For coherent theories, coming up with examples requires a bit more algebraic groundwork. Up to Morita equivalence, a coherent propositional theory is the same thing as the theory of prime filters of a (unique up to isomorphism) distributive lattice (see Johnstone, 2002, Remark D1.4.14). Given any such distributive lattice B the category of models of the corresponding theory can then be identified with the spectrum of B . And if B is Boolean (as a lattice) then its spectrum will be discrete. This means that up to equivalence of their categories of \mathcal{S} -models we can only recover a coherent propositional theory up to the cardinality of its spectrum. However, there are many examples of non-isomorphic Boolean lattices whose spectra have equal cardinalities. Indeed a similar idea as our previous example works again here. Let B_1 be the Boolean algebra generated by a countably infinite number of elements p_0, p_1, \dots and let B_2 be the Boolean algebra generated by the same elements plus the relation $p_0 \leq p_i$ for all $i \in \mathbb{N}$. B_1 is atomless whereas B_2 has an atom

and therefore B_1 and B_2 cannot be isomorphic. However, the cardinality of both their spectrums is equal to 2^{\aleph_0} a fact which can be seen by noting that homomorphisms $B_1 \rightarrow \mathbf{2}$ correspond exactly to homomorphisms $\phi: B_2 \rightarrow \mathbf{2}$ such that $\phi(p_0) = 0$. \lrcorner

The lesson here is that the semantic category of a theory — i.e. the models of that theory, and homomorphisms between models — generally contains less information than the theory itself does. A fortiori, the class of models of a theory contains less information than the theory itself does. (And this is what’s wrong with the original semantic view of theories.)

It is completely natural to ask: if $\text{Mod}(T)$ does *not* contain the same amount of information as T , then what information or structure must be added to $\text{Mod}(T)$ in order to recover T ? Think of the question this way: $\text{Mod}(T)$ is a collection of models and arrows between models, including automorphisms (i.e. arrows from a model to itself). What other information about $\text{Mod}(T)$ can be extracted from the theory T ?

A classic answer to this question was given (for the propositional case) by Marshall Stone. Stone noted that T implicitly contains topological information about $\text{Mod}(T)$. In particular, let’s say that a sequence m_1, m_2, m_3, \dots of models in $\text{Mod}(T)$ converges to a model m_0 just in case for any sentence ϕ of Σ , the truth value $m_i(\phi)$ is eventually equal to $m_0(\phi)$. This notion of convergence defines a topology on $\text{Mod}(T)$. Letting $\underline{\text{Mod}}(T)$ denote the corresponding topological space, Stone’s duality theorem establishes the following:

Theorem (Stone Duality). *The collection of compact open subsets of $\underline{\text{Mod}}(T)$ forms a Boolean lattice that is equivalent, as a category, to C_T .*

In other words, from $\underline{\text{Mod}}(T)$ we can reconstruct T up to its syntactic category, i.e. up to (SE). Thus, in the case of propositional theories, the *topological* semantic category $\underline{\text{Mod}}(T)$ contains as much information as T .

But what now about the case of predicate logic? Here the situation is complicated by the fact that there are typically many non-trivial arrows between models. Can the category $\text{Mod}(T)$ still be supplemented with topological information in order to recover T ? The answer here is: Yes, sort of. Although there is still no result that perfectly generalizes Stone Duality, some important partial results have been obtained by Makkai (1991) and Awodey and Forssell (2013).

Of course, philosophers of science should be eager to understand these duality results because of the important lesson they teach about the collection of models of a theory:

The category of models $\text{Mod}(T)$ of a theory T does not generally contain all the information that is contained in the original theory T . The content of T might include information, e.g., about topological relations between models.

Let's rephrase that moral one more time, now trying to make it absolutely clear as a friendly amendment to the semantic view of theories:

The mathematical content of a scientific theory T is not exhausted by the class of models of T .

Obviously, this moral from first-order logic doesn't generalize directly to scientific theories in the wild (e.g. classical mechanics, general relativity, quantum mechanics). However, the results from first-order categorical logic strongly suggest that in the case of scientific theories, we would similarly go wrong if we identified the mathematical content of a theory T with its category $\text{Mod}(T)$ of models⁸ — for the theory might make use of further structures on $\text{Mod}(T)$, perhaps topological (as in general relativity), or measure-theoretic (as in statistical mechanics), or perhaps some sort of monoidal or tensor structure (as in quantum mechanics and quantum field theory). For some evidence for this claim, with reference to specific scientific theories, see (Curiel, 2014) or (Fletcher, 2015), and for some related discussion see (Lal and Teh, 2015).

1.4 On Methodology

A few more remarks are now in order concerning the big-picture methodology that we are envisioning and category theory's role in it. We do not want to fall into the trap — all too common in 20th century philosophy — of being blinded by the glow of a shiny new piece of formal apparatus. We are fully aware that, from a mathematical point of view, a syntactic category is regarded merely as

⁸An important clarification: when we say here that we would go wrong if we identified the mathematical content of a theory T with its semantic category $\text{Mod}(T)$, recall that this means the semantic category *over the category of sets* \mathcal{S} . So what we are saying here is that a theory cannot be recovered up to (ME) from its category of models in \mathcal{S} . Nevertheless, conceptual completeness for coherent logic (see Johnstone (2002, Theorem D3.5.9)) says that this is very close to being true: equivalence of the semantic categories of T and T' does imply (ME) as long as this equivalence is induced by an interpretation I of T into T' at the level of syntax (i.e. a coherent functor $I: P_T \rightarrow P_{T'}$). Relatedly, there is another sense in which a theory T is actually recoverable up to (ME) from its semantic category: if we consider semantic categories over arbitrary Grothendieck toposes \mathcal{E} and stipulate that the equivalence $\mathcal{E}\text{-Mod}(T) \simeq \mathcal{E}\text{-Mod}(T')$ is *natural* in \mathcal{E} , then from this alone we can conclude that T and T' are Morita equivalent. This means that the extra structure that we need to place on $\text{Mod}(T)$ (understood as the category of \mathcal{S} -models) in order to recover T up to Morita equivalence corresponds exactly to the requirement of naturality in the class of Grothendieck toposes.

a technical device useful for proving other results, e.g. completeness theorems.⁹ Yet we believe that the conceptual value of the kind of formal apparatus best encapsulated by the syntactic category of a theory goes beyond its practical use *within* mathematics. The syntactic category is a formal construction realizing a possible synthesis of an opposition that — under different guises — has occupied philosophy throughout its history: that between syntax and semantics. This opposition has proved especially vexing in the philosophy of science as we saw in the above-summarized debate between the syntactic and the semantic views of theories. Category theory provides a resolution of the syntax/semantics opposition (understood formally) in the form of the syntactic category.¹⁰ It would be shallow at best and narrow-minded at worst for the philosophy of science not to attempt to draw lessons from category theory’s successes in this manner. To be sure, these successes take place in a very “sterilized” mathematical setting dealing with first-order theories that have little in common with the wild, unaxiomatized mathematical beasts that physicists contend with. Unlike W.v.O. Quine, we have no ideological commitment to regimenting theories in first-order logic. Rather, we see first-order logic as providing a manageable testing ground for more general ideas about theoretical structure.

What we are saying here clearly amounts to a methodology based on an analogy between the categorical metamathematics of first-order theories and the philosophy of scientific theories.¹¹ Category theory brings to the table new constructions and concepts with which to study the metamathematics of first-order theories. We ask: can these concepts carry over to the philosophy of science in any fruitful sense? Put less modestly: can category-theoretic thinking about first-order theories re-invigorate (perhaps even overhaul) philosophical thinking about scientific theories? We are convinced that it can.

All this raises some broader issues on the very relationship between formal mathematical work and philosophy. Let us make a few brief remarks here on how we envision this relationship (which may betray a not entirely uncontroversial view of philosophical methodology). At its best, philosophy advances by appropriating new vocabularies (or “concepts”) and deploying them to old tasks. It is certainly constrained in this process of appropriation by some kind of universal (or “transcend-

⁹And this extends to syntax very far removed from first-order logic, e.g. the metamathematics of simple type theories or of Martin-Löf type theories are also studied via structure-bearing syntactic categories.

¹⁰Inspired, it has to be said, by Lindenbaum-Tarski-style constructions that predate it.

¹¹If one finds oneself in a particularly optimistic mood, one might hope that the analogy carries over even to the study of theories of metaphysics, i.e. to what now goes under the name of metametaphysics.

dental”) logic, i.e. a universal *form* that constrains this process in just the right way to guarantee its *meaning*. Other than that, however, we believe that every realm of discourse is fair game for philosophical appropriation — and this is especially true of mathematics and the natural sciences. Contrary to how this dictum has been interpreted across large swathes of analytic philosophy, this does not mean, for us, that the assertions of mathematics and the natural sciences are to be taken as the immovable *datum* around which philosophy must somehow build a niche for itself. Rather, thorough knowledge of mathematics and the natural sciences are fruitful for philosophy not so that we know *what* these scientists *assert* but rather *how* these scientists *think* and especially *what kind* of thinking has proved fruitful and useful to them. Philosophy is not to be reduced to formal work; rather, formal work is to provide the canvas on which philosophy is to be painted. It is in this spirit that we take formal work to be essential to the philosophical enterprise.

We believe that category-theoretic thinking and category theory are ripe for philosophical appropriation (in the above-described spirit). As clarified in the beginning, this does not mean that we are claiming that category theory *is* the right way to think about this or that phenomenon. Our hope, rather, is that category theory will allow us to see old problems under a new light and hopefully give rise to new forms of thinking in the process. Initially this process may appear as working by (unfounded) analogy, viz. that between the categorical metamathematics of first-order theories and the philosophy of scientific theories. Surely, one might object, this analogy requires substantiation. Otherwise, what is there to stop one from accusing us (with a hint of irony) of making a category mistake?

To this we have two things to say. Firstly, our formal work is constrained by *mathematical* criteria: it is – one might say – mathematical work done with philosophical goals in mind. This provides some minimal protection from nonsense and contradiction even if the analogy we purport to rely on proves shaky or unconvincing. Secondly, we believe that the burden of proof is not with us. It is our accuser who would have to provide some proof that formal work in mathematics is altogether unsuitable to serve as a philosophical canvas. The bond between mathematics and philosophy has always been strong (at least in the modern era). Leibniz, for example, thought that he could rely on the new science of the analysis of infinitessimals as a conceptual guide to his sought-after *characteristica universalis* and Leibniz was certainly no fool. In a Leibnizian (and therefore overly optimistic) spirit we similarly feel confident in taking our conceptual cues from

mathematics and especially – for the present purposes – from category theory.

But, as a reminder, this does not mean that we take the mathematics of category theory to *reveal* some essential truth about scientific theories. Nor do we take category theory to help us in the process of extracting from the viewpoint of theories-as-categories a sharper vision of what it means to be a scientific theory. That is to say, we do not see the virtue of the categorical approach to be that it adds one more layer of differentiation to all previously existing views (e.g. the naive semantic view, the Carnapian view etc.) such that, in abstracting the common features of all of them, we get an even sharper picture of the “essential nature” of a scientific theory. Rather, for us, category theory adds one more layer of differentiation to our thinking about scientific theories – and in doing so adds one more piece to the great unfolding puzzle of the exact relationship between natural science and the reality it describes.¹²

As much respect as philosophy should accord mathematicians and scientists – and philosophers should certainly try their best to understand the work of mathematicians and scientists – this does not mean that philosophy should be reduced to compiling grocery lists to sate the ontological appetites of science. Philosophers are engaged in their own creative work which feeds off from mathematics and natural science (and many, if not all, disciplines) but is not governed by them. A scientific approach to philosophy does not – in our mind – mean that philosophy is to be reduced to scientific exposition. This is the spirit in which we urge philosophers – and especially philosophers of science – to engage with category theory.

In summary, our work does not rely on the relationship between the philosophy of science and the categorical metamathematics of first-order theories because, somehow, first-order theories are to be thought of as adequate “toy models” of scientific theories. This may very well be the case – and we certainly believe it to be the case – but our methodology does not rise or fall based on this claim of adequacy. Rather, what we find in (not exclusively categorical) metamathematics is the emergence of certain *concepts*, *relations* and *oppositions* that very much reflect those that the philosophy of science also engages with. For example: the notion of a theory, the way it relates to syntax, the way it relates to semantics, and the manner in which syntax and semantics are op-

¹²The manner in which we find the appropriation of formal work by philosophy fruitful can perhaps best be summarized in Cassirer’s attitude to language: “the true universal ‘essence’ of language [is] no longer sought in abstraction from differentiation, but in the *totality* of differentiations.” (Cassirer, 1953, p. 155) (To be perfectly precise, Cassirer is here summarizing Humboldt’s views on language – but it is clear that he too endorses such a view.)

posed. Both philosophy of science and (not exclusively categorical) metamathematics deal with this situation (where “theory”, “syntax” etc. are understood differently). In (exclusively) categorical mathematics, the opposition between syntax and semantics is resolved through the construction of the syntactic category. So the fact that there should be something like the “syntactic category” of a scientific theory does not stem from a conviction in some essential *similarity* between first-order theories and scientific theories. It stems, if anything, from the empirical observation that the language and the concepts surrounding first-order theories in how they are studied metamathematically (syntax, semantics etc.) is similar to the language and concepts surrounding scientific theories in how they have been studied in the philosophy of science (at least from Carnap onwards). It is this observation that grounds our conviction in the fruitfulness of the above-described appropriation of the categorical vocabulary – and it is this conviction that provides the impetus for the kind of technical projects that we outline in the following section.

2 The category of theories

Let us now put the transcendental justification of our methodology to the side and return to more practical questions: why should philosophers of science invest the time and effort in learning category theory? What good is it to think of theories as categories? We claim that the primary virtue of this approach is that it allows us to see theories themselves as the objects of a category: *the category of theories*. We can then apply the tools of category theory to understanding the structure of this larger category, how individual theories sit within it, and how theories are related to each other.

As before, our initial focus is on the case of theories in first order logic. As described in the previous section, each first-order theory T corresponds to a syntactic category C_T (which we could also take to be P_T). We let the collection of all such C_T be the objects of a category \mathbf{Th} , the category of all first-order theories.

Again, we have some fine-grained control over the definition of the category \mathbf{Th} of theories. The main possibilities for \mathbf{Th} are as follows:

- **Coh** the category of coherent categories (i.e. syntactic categories of coherent theories);
- **dCoh** the category of decidable coherent categories (cf. Awodey and Forssell, 2013);

- **BCoh** the category of Boolean coherent categories (i.e. syntactic categories of first-order theories over classical logic);
- **Pretop** the category of pretoposes.

These categories are arranged roughly as follows:

$$\mathbf{BCoh} \subseteq \mathbf{dCoh} \subseteq \mathbf{Coh}$$

and

$$\mathbf{Pretop} \subseteq \mathbf{Coh}$$

where the subset symbol indicates a full inclusion of categories.

Choosing our “category of theories” from among these (or other) options clearly amounts to choosing a notion of equivalence for our theories. Namely – as is usual in categorical thinking – choosing where your objects of study live automatically determines what it means for your objects of study to be “isomorphic”. This is an important point to keep in mind: choosing a notion of equivalence for theories and choosing a “category of theories” are not two independent choices. Choosing one determines the other – there is only one degree of freedom here.

In our opinion, the two most natural choices for **Th** are **Coh** or **Pretop**. This shouldn’t come as a surprise: we’ve already said that the two notions of equivalence that interest us the most are (SE) and (ME) and these correspond exactly to choosing **Coh** (for (SE)) and **Pretop** (for (ME)) as our preferred categories of theories. This is because (SE) identifies theories with their syntactic categories and every coherent category is the syntactic category of some coherent theory, whereas (ME) identifies theories with the pretopos completion of their syntactic categories and every pretopos can be seen to arise in this manner (although, of course, inequivalent coherent categories may have equivalent pretopos completions). As to restricting ourselves to coherent logic recall, in particular, the arguments of Makkai and Reyes to the effect that coherent logic is to be preferred to full first-order logic.

Furthermore there is a natural relation between these two categories: for each coherent category C , there is a unique pretopos $P(C)$ and functor $\eta_C : C \rightarrow P(C)$ satisfying a suitable universal property. In short: every coherent category has a unique pretopos completion — an operation corresponding roughly to taking a “maximal Morita extension” of the original theory.

In fact, $P : \mathbf{Coh} \rightarrow \mathbf{Pretop}$ is a 2-functor (see Makkai, 1987). And even more is true: \mathbf{Pretop} is “almost” a reflective sub-2-category of \mathbf{Coh} . More precisely, \mathbf{Pretop} is a full reflective sub-2-category of \mathbf{FinSit} , the 2-category of “finitary” sites (i.e. categories equipped with finitely-generated Grothendieck topologies) – this is a special case of a far more general result proven by Shulman (2012).

For the purposes of this paper, we needn’t make a decision about the precise definition of \mathbf{Th} . However, for concreteness, let us say that we are in favor of the identification $\mathbf{Th} = \mathbf{Pretop}$. (As explained above, this means that we are effectively choosing (ME) as our preferred notion of equivalence.) Although nothing we say here hinges on this choice, let us say a couple of things about why it is a natural choice to make (other than our faith in (ME)). Firstly, there is a very precise sense in which – from a logical point of view – the pretopos completion of a coherent category adds only those concepts that are already *definable* from the coherent structure of the original category. Secondly, the pretopos completion is the *maximal* such extension, i.e. it contains *everything* that is definable from the coherent structure of a coherent category (for a precisification of this statement see (Harnik, 2011)). As such, if one agrees with us that the initial syntactic presentation of a theory does not constitute its essential content then moving from \mathbf{Coh} to \mathbf{Pretop} should seem a very reasonable move to make.

Now what are the arrows in \mathbf{Pretop} ? Since the objects of \mathbf{Pretop} are categories, the arrows should be functors. Perhaps surprisingly, the arrows we care about in this particular case are obtained by considering a pretopos as a coherent category (recall that every pretopos is coherent). But do coherent categories have additional structure that ought to be preserved by our arrows? The answer, in short, is yes: a coherent category has limits and colimits that encode various syntactic structures — in particular, conjunction, disjunction, and existential quantification. Thus, we define an arrow between pretoposes P and P' to be a *coherent functor* in the sense of (Johnstone, 2002, p. 34), also called a *logical functor* in (Makkai and Reyes, 1977, p. 121). In short, we consider \mathbf{Pretop} as a full subcategory of \mathbf{Coh} .¹³

¹³The fact that it is reasonable to do so essentially boils down to the above-mentioned fact that the pretopos completion of a coherent category is “definable” (in a precise sense) from the coherent structure of the category in question. In particular, even though coherent morphisms will not, in general, preserve arbitrary coproducts or coequalizers, they will preserve binary coproducts and coequalizers arising from equivalence relations. This means that “disjoint unions” and “quotients by equivalence relations” are concepts within the grasp of coherent logic and taking the pretopos completion of a coherent category amounts to a (maximal) “definitional extension” of the original coherent category by these definable concepts. Indeed, removing the scare quotes from the previous sentences and making this way of talking about pretopos completions fully precise was one of the motivations behind (Barrett and

Choosing coherent functors as arrows has the nice consequence that arrows from P_T to P_S correspond to translations of the theory T into S (so do, incidentally, arrows from C_T to C_S). A way to see this is the following: each P_T contains the so-called *generic model* M_T of T (Johnstone, 2002, Proposition D1.4.12.(ii)). This is a model of T taken in the pretopos completion P_T (recall that P_T is coherent and therefore has the capacity to model any coherent theory T). It is *generic* in the sense that it satisfies exactly those sentences that are provable in T , i.e.

$$M_T \models \phi \iff T \vdash \phi.$$

Now, since coherent functors preserve the coherent structure and since models of coherent theories in coherent categories are built using (only) that coherent structure, we have that any coherent functor $F: P_T \rightarrow D$ into a coherent category D will give us a model $F(M_T)$ of T in D . In particular, when $D = P_{T'}$ for some other theory T' then $F(M_T)$ is a model of T in $P_{T'}$. But a (\mathcal{S} -)model of T' is simply a coherent functor from $P_{T'}$ into \mathcal{S} . Therefore, any model $G: P_{T'} \rightarrow \mathcal{S}$ of T' will also give rise to a model of T , viz. $GF: P_T \rightarrow \mathcal{S}$. In plain terms: any model of T' contains a model of T ; this is just another way of saying that there is a translation of T into T' . Indeed, in (Pitts, 1989) translations of a theory T into another theory T' are *defined* to be models of T in $P_{T'}$.¹⁴ Finally, it is important to note that everything we've said in this paragraph can be said pretty much verbatim for syntactic categories themselves (rather than their pretopos completions). More on translations and definability at the level of syntactic categories can be found in (Caramello, 2012).

It should immediately be pointed out that **Pretop** is most naturally thought of as a 2-category, rather than just a category. Recall that a 2-category \mathbf{C} is (roughly speaking) a category such that for any two objects a, b of \mathbf{C} , instead of $\mathbf{C}(a, b)$ being a *set* of arrows from a to b , it is a *category*; and the composition operation on arrows is functorial (see Borceux, 1994; Lack, 2010). The arrows in the category $\mathbf{C}(a, b)$ are called *2-cells*. The paradigm example of a 2-category is **Cat**, the category of (small) categories, with functors as arrows, and natural transformations as 2-cells. That is, if C and D are categories, then **Cat**(C, D) is the category whose objects are functors $F: C \rightarrow D$, and whose arrows are natural transformations between such functors. Similarly, we define **Pretop**(P, P') to be the category whose objects are coherent functors from P to P' , and

Halvorson, 2015b) and (Tsementzis, 2015).

¹⁴Which is the same thing as saying that translations are functors $C_T \rightarrow P_{T'}$ which in turn is the same thing as saying that translations are functors $P_T \rightarrow P_{T'}$.

whose arrows (2-cells) are natural transformations.¹⁵

Basking in the full 2-categorical glory of **Pretop** is no mere pretension, nor is it a pointless exercise to prepare us for the altitude sickness that comes with the ever steepening ascent towards higher category theory. It is, rather, a perspective that brings, among other things, new purely logical insights. To convince oneself of this one need look no further than Pitts’ proof of conceptual completeness for coherent logic (resp. intuitionistic logic) using the 2-categorical structure of **Pretop** (resp. **HPretop**, the category of Heyting pretoposes) – see (Pitts, 1987, 1989).

Given this elegant formal framework, numerous technical questions — of philosophical interest — suggest themselves. We will now list and discuss a few such questions.

1. We saw that if the syntactic categories C_T and C_S are equivalent, then the theories T and S are Morita equivalent, i.e. that (SE) implies (ME). We also saw that the converse fails: two theories can be Morita equivalent even though their syntactic categories are inequivalent.

An interesting question we can now raise is this: What kind of conditions can we place on theories such that (SE) coincides with (ME)? (One well-known case in which they do coincide is that of algebraic theories.) Furthermore, what can we say about Morita equivalences over particular categories? For example, for any two first-order theories T and T' whose semantic categories are equivalent (over \mathcal{S} , and not necessarily naturally) how can we characterize their relation from a purely syntactic standpoint? And what about their thin categories? Namely, if

$$\mathcal{S}\text{-Mod}_e(T) \simeq \mathcal{S}\text{-Mod}_e(T')$$

then – purely syntactically – how are T and T' related?

2. Let’s consider some natural relations between theories. First, let T be a theory in signature Σ , and let T' be an extension of T by some additional axioms, also in the language Σ (such “extensions” are also called “quotients”). Then there will be a canonical functor $F : C_T \rightarrow C_{T'}$. What can we say about this functor? What features does it have? In the case of geometric theories a lot of work in this area has been done by O. Caramello starting with the “Duality Theorem” in her Phd thesis – for a big picture view see (Caramello, 2009, 2010). See also (Forssell, 2013).

¹⁵A 2-category is a “strict” version of a bicategory in the sense of (Bénabou, 1967).

Similarly, let T be a theory in signature Σ , and let T' be the same theory, but considered in a larger signature Σ' . Again there will be a canonical functor $F : C_T \rightarrow C_{T'}$. What features does this functor have?

Furthermore, philosophers of science have been interested in questions about when one theory T' is *reducible* to another theory T . Can reduction be thought of as a functor $F : C_{T'} \rightarrow C_T$ (see Van Benthem and Pearce, 1984)? How does this functorial account compare to the classical Nagelian account (see Nagel, 1979, Ch 11)? How does this functorial account compare to semantic accounts (see Bickle, 1998)?

3. If we were to think of **Coh** or **Pretop** as merely a category, then a natural technical question might be: does this category have limits? Or, does this category have colimits? And, if it does have limits or colimits, then do these have any sort of natural interpretation as operations on theories? For example, is there any sense in which a coproduct of two coherent categories represents a sort of amalgamation of the two theories? And if so, does the notion of an “amalgamation of theories” have a clear interpretation?

However, it’s more natural to think of either **Coh** or **Pretop** as a 2-category, in which case the better questions have to do with the existence of limits and colimits in the bicategorical sense. Does **Pretop** have 2-limits and 2-colimits? And if it does, do these limits and colimits have a natural interpretation as operations on theories?¹⁶

Philosophers of science should be particularly interested in whether the categorical structure of **Pretop** can be used to explicate various relations between theories, such as limiting relations.

4. The relation between “thick” and “thin” semantic categories provides very fertile ground for investigation, as noted above. One obstruction here is that the thin semantic categories of first-order theories are almost invariably “too thin” in the sense that they rarely contain interesting categorical structure (e.g. limits, colimits etc.)

On the other hand, it is perhaps worth investigating relations between pairs of theories (T_1, T_2) such that there are interesting functors

$$F: \text{Mod}_e(T_1) \rightarrow \text{Mod}(T_2)$$

¹⁶Makkai (1995) proves a version of the Craig interpolation theorem using the 2-categorical version of a pushout of syntactic categories.

where T_1 would be thought of as the “background spacetime” theory and T_2 would be thought of as the “physical system theory”. To be a little more specific, this formal situation seems to us an interesting generalization of the nowadays very common situation (e.g. with (Q)FTs) where we have a category of “state spaces” represented by algebraic objects (e.g. Hilbert spaces, C^* -algebras) and a category of “spacetimes” understood as categories of “physical spacetimes with embeddings as arrows” and where a theory of physics is defined as a functor relating those two.

For example, Fewster (2015) defines a locally covariant theory to be a functor

$$F: \mathbf{Bkgnd} \rightarrow \mathbf{Phys}.$$

In Fewster’s set-up, \mathbf{Bkgnd} seems to us to be best understood as essentially the thin category of models of some theory, since the morphisms are basically elementary embeddings (see Fewster, 2015, p. 4). On the other hand, the categories which he calls \mathbf{Phys} seem to us to be best understood as the thick categories of models of some theory, since they are usually categories of algebraic structures.

So perhaps it’s worth looking at what kind of interesting things can be said about such functors, when T_1 and T_2 are first-order theories: what can we say about them, modulo some constraints on T_1 , T_2 and F ? Is this a fruitful general set-up with which to study properties of (Q)FTs?

5. The (2-)category \mathbf{Pretop} might be used to explicate the notion of a *symmetry* of a theory.

In recent literature in philosophy of science, discussion of symmetries has not been technically well controlled. Motivated perhaps by the semantic view of theories (see more below), these discussions typically suppose that a symmetry operates on the class of models, or on the set of solutions to an equation — with almost no attention to formal constraints on such mappings. We should be clear, however, that there are both syntactic and semantic notions of symmetries — as can be made precise with the category \mathbf{Th} of theories (understood here as \mathbf{Pretop}). Let T be a theory and let P_T be its syntactic category, which is an object of \mathbf{Th} .¹⁷ Then we propose:

¹⁷A note of clarification on terminology: we are now using the term “syntactic category” to refer to an arbitrary

A syntactic symmetry of T is an auto-equivalence of the syntactic category P_T .

Recall that a coherent functor $F : P_T \rightarrow P_S$ corresponds to a translation from T into S . Thus, an auto-equivalence $F : P_T \rightarrow P_T$ corresponds to a translation of T into itself (i.e. a sort of permutation of the vocabulary of T).

But more is true. A coherent functor $F : P_T \rightarrow P_S$ induces a functor $F^* : \text{Mod}(S) \rightarrow \text{Mod}(T)$ from models of S to models of T (see Gajda et al., 1987; Makkai and Reyes, 1977). In particular, an automorphism $F : P_T \rightarrow P_T$ induces a functor $F^* : \text{Mod}(T) \rightarrow \text{Mod}(T)$ on the category of models of T , which raises another technical question:

Is it true that any essentially invertible functor $G : \text{Mod}(T) \rightarrow \text{Mod}(T)$ has the feature that $G = F^$ for some coherent functor $F : P_T \rightarrow P_T$?*

The answer to this question is No, as can be seen by again looking at the example of the two propositional theories. Thus, philosophers of science should *not* necessarily suppose that any auto-equivalence $G : \text{Mod}(T) \rightarrow \text{Mod}(T)$ should count as a symmetry of T . But can we say something about further conditions on G so that it is indeed dual to some functor F on the syntactic category P_T ?

6. Since the demise of the syntactic view of theories, philosophers of science have been fond of pointing out that interesting scientific theories — even those in rigorous mathematical physics — typically fail to admit a first-order axiomatization. Thus, we might conclude that a typical scientific theory cannot be described by an object in **Coh** or **Pretop**. We already dealt with this point in Section 1.4 from a more abstract perspective. Let us now add a few more words. Firstly, is this dismissal too fast? Note that some logics stronger than first-order logic — e.g. geometric logic, or even higher-order logics — will also give rise to syntactic categories that are coherent. In fact, if a logic is stronger than first-order logic, then the corresponding categories can be expected to have more structure than coherent categories. For example, in the case of intuitionistic type theory (ITT), there is a syntactic category C_{ITT} (produced by a similar but different process than the one we outlined for first-order theories T) which bears the structure of an elementary topos (see (Lambek and Scott, 1986) for the classical account).

object of the category of theories **Th** rather than the explicit construction as carried out in Section 1.1.

Relatedly, recent work in “cohesive” homotopy type theory (CoHoTT) uses syntactic methods (based on a logic much more “exotic” than first-order logic) to study higher gauge theories (see (Schreiber, 2013) as well as (Corfield, 2016) in this volume). Among other things, Schreiber envisions a certain class of ∞ -toposes (the so-called “cohesive” ones) as the correct setting at which to study the foundations of higher gauge theories, *exactly because* these cohesive ∞ -toposes are (conjectured to be) the syntactic categories of CoHoTT. Of course, Schreiber’s use of syntactic methods is not motivated by considerations on theoretical equivalence and the structure of scientific theories – his mathematical work is carried out with the explicit goal of articulating a general foundation for higher gauge theories. Nevertheless, the way in which he blends syntactic and semantic methods is a great illustration, in our opinion, of how representing scientific theories as categories (in his case *higher* categories) is an illuminating perspective to take.

So, is being a coherent category a *minimal* necessary condition for representing a bona fide scientific theory? In any case, it would be natural to ask:

Given a formalized theory T of the empirical sciences, is there some category C_T that can be thought of as the “syntactic category” of T ?

Consider a couple of examples. First, let T be Einstein’s general theory of relativity. We do have some sense of what the *semantic* category — i.e. the category of models of T — ought to be, viz. the category of differentiable manifolds with Lorentzian metric and stress-energy tensor satisfying Einstein’s field equations. But is there a category C_T that could be considered the syntactic category of GTR?¹⁸ And could the failure to distinguish between semantic and syntactic points of view be partially responsible for some of the difficulties that philosophers have had understanding the nature of symmetries in GTR?

As a second example, let T be quantum mechanics. In this case we also have a sense of what the semantic category of T ought to be — namely, the category of (finite-dimensional) Hilbert spaces and linear operators. (This example was one of the main motivators for the semantic view, at least in the mind of van Fraassen.) Now, what might the syntactic category C_T of T look like? Is there a way to present quantum mechanics syntactically? Should we expect C_T

¹⁸One might hope for some help here from investigations in synthetic differential geometry. See, for example, (Reyes, 2009).

to be something like a symmetric monoidal category instead of a coherent category? (Perhaps some help here might come from recent work in categorical quantum logic, see e.g. Coecke and Kissinger, 2016.) Note, finally, the importance of finding a syntactic presentation of quantum mechanics: following Quine’s dictum that, “to be is to be the value of a variable,” to find a syntactic presentation of quantum mechanics is tantamount to explicating the ontology of that theory.

7. One of the complaints that van Fraassen — among several others — raised against the syntactic view of theories was that it couldn’t make sense of the notions of empirical adequacy and empirical equivalence (see Van Fraassen, 1980). In contrast, it was claimed that a semantic approach to theoretical structure provided the resources to explicate these notions.

To be honest, we doubt the claim that the semantic view has an advantage in this regard. However, the categorical approach provides new insight into the structure of theories. Might it be useful for articulating these notions that are so important to an empiricist philosophy of science?

Recall that the simplistic method of isolating empirical content of a theory T ran as such: given the signature Σ , suppose that $\Sigma = \Sigma_t \cup \Sigma_o$, where Σ_t contains theoretical terms, and Σ_o contains observation terms. Then the empirical content of T is simply $T|_{\Sigma_o}$, i.e. the consequences of T in the subvocabulary Σ_o .

This method of isolating empirical content leads to absurdities, as noted by Achinstein, Putnam, van Fraassen and others. But can some version of it be revived by thinking in terms of a syntactic category? In particular, suppose that C is a coherent category, representing some theory T . Now suppose that $C = C_t \sqcup C_o$, i.e. C is the coproduct of two other coherent categories.¹⁹ Might this sort of division be a way of representing the empirical content of C ? Could it overcome the objections that were leveled by the critics of the syntactic view of theories?

Or perhaps there is some other way of representing empirical content? Perhaps an empirical theory should be thought of as a pair of categories (C, D) , where C is the syntactic category (as before), and D is the “category of predictions”; and perhaps some relation between C

¹⁹Technically, the more appropriate notion here is a 2-coproduct.

and D should be required.

But the most important test of these categorical ideas will be whether they can make sense of *actual* empirical theories. On the one hand, the simplistic method of isolating empirical content — via writing a signature as $\Sigma = \Sigma_t \cup \Sigma_o$ — is useless when applied to all but the most trivial empirical theories. For example, even a theory as mathematically rigorous as General Relativity doesn't have an obvious “language,” and so it's not clear how to represent its empirical content syntactically. On the other hand, the semantic approach to empirical content is so vague that no results of any interest have been proven about it (or at least, not to our knowledge). We should hope that a categorical approach to these issues will both be flexible enough to apply to actual theories, and will be precise enough to allow interesting results to be proven.

3 On the duality of syntax and semantics

Recall that the logical positivists hoped to provide an explication for the notion of a *scientific theory*.²⁰ That is, they hoped to be able to say that a scientific theory is a certain sort of (rigorously defined) mathematical object. But what kind of object? According to the earliest proposals (by Carnap and others), a theory is a set of sentences in a formal language. This proposal and its later elaborations have come to be known as the *syntactic view of theories*.

As is well known, the syntactic view of theories was subjected to severe criticism in the later 20th century. The consensus in the 1970s was the the syntactic view couldn't be salvaged, and required a wholesale replacement. The proposed replacement was the so-called *semantic view of theories*, which claims that a scientific theory is a collection of models — perhaps the models of some first-order logical theory, or perhaps a collection of models of some more general sort.²¹

It has long been thought that the semantic view of theories has many advantages over the syntactic views — see e.g. the works of Suppe, van Fraassen, and Lloyd, et al.. Of course, that claim presupposes that there is a genuine dilemma of choice between the two points of view. Only a couple of isolated philosophers have suggested that this might be a false dilemma (see e.g. Friedman,

²⁰For further elaboration of this story, see (Halvorson, 2015).

²¹To be clear, Van Fraassen (2014) has recently pointed out that for him, a theory is a class of models together with representational content. However, for this discussion, we are concerned only with the mathematically representable part of a theory.

1982). In this section, we survey mathematical results that argue for a formal duality between the syntactic and semantic points of view. We also propose that this duality could be exploited in order to better understand the structure of scientific theories.

Recall the previous discussion of Stone duality for theories in propositional logic. While it's not true that a propositional theory can be reconstructed from its category of models alone (i.e. the set of ultrafilters on the Lindenbaum algebra), it can be reconstructed from the category of models *plus* relevant topological information.

In recent years, logicians have attempted to generalize Stone duality to the case of full first-order logic. And while the results to date are only partial, they all point in a similar direction.²²

First, Makkai (1991) makes use of an insight from Los' theorem: if $\{m_i\}_{i \in I}$ are models of a theory T , then so is an ultraproduct $\prod_{i \in I} m_i / \mathcal{U}$, where \mathcal{U} is an ultrafilter on I . What's more, in the case where T is a propositional theory, the ultraproduct $\prod_{i \in I} m_i / \mathcal{U}$ is simply the Stone topology limit of the sequence $\{m_i\}_{i \in I}$ along the ultrafilter \mathcal{U} . In other words, in the propositional case the Stone topology on $\text{Mod}(T)$ can be alternatively described as "ultraproduct structure" on $\text{Mod}(T)$; and the relevant functors $F : \text{Mod}(T) \rightarrow \text{Mod}(T')$ are those that preserve this ultraproduct structure (i.e. that are continuous in the Stone topology).

Now Makkai defines an *ultracategory* to be a category with a sort of ultraproduct structure (see Makkai, 1991). Of course, the motivating example of an ultracategory is the category $\underline{\text{Mod}}(T)$ of models of a first-order theory. Then the question arises:

Can a theory T be reconstructed from the corresponding ultracategory $\underline{\text{Mod}}(T)$?

Makkai shows that the answer is Yes. For any pretopos P , let $\Theta(P) = \mathbf{Coh}(P, \mathcal{S})$ denote the category of coherent functors from P into the category \mathcal{S} of sets (i.e. models of the theory corresponding to P). Makkai shows that there is another functor $\Gamma : \mathbf{UCat} \rightarrow \mathbf{Pretop}^{op}$, such that $(\Gamma \circ \Theta)(P)$ is equivalent to P . Stated more generally: there is a pair of adjoint functors as follows:

$$\begin{array}{ccc} & \Theta & \\ & \curvearrowright & \\ \mathbf{Pretop}^{op} & & \mathbf{UCat} \\ & \curvearrowleft & \\ & \Gamma & \end{array}$$

²²Such dualities as the ones we will outline below sometimes go under the name of "Isbell Duality" and many examples have been studied at a very high level of generality – see Porst and Tholen (1991) and Barr et al. (2008) for a sampling.

Unfortunately, this adjunction is *not* an equivalence of categories, as in the case of propositional theories (i.e. Stone duality for Boolean algebras). In particular, not every ultracategory is of the form $\text{Mod}(T) \simeq \mathbf{Coh}(P_T, \mathcal{S})$, for some first-order theory T . In slogan form: there are more ultracategories than there are coherent theories.

A more recent attempt to generalize Stone duality has been undertaken by Awodey and Forssell (2013). Here the insight comes not from model theory (as in the case of Makkai’s ultraproducts), but from topos theory. Recall that Joyal and Tierney (1984) proved that for every Grothendieck topos \mathcal{E} , there is a localic groupoid G such that $\mathcal{E} \simeq \mathbf{B}(G)$, where $\mathbf{B}(G)$ is the topos of continuous actions of G . It was also shown by Butz and Moerdijk (1998) that when \mathcal{E} has enough points — as is the case when $\mathcal{E} \simeq \mathcal{E}_T$ is the classifying topos of a coherent theory — then G may be taken to be a topological groupoid.²³

The models of a (coherent) theory T naturally form a category $\text{Mod}(T)$. Now, if we eliminate all non-isomorphism arrows from $\text{Mod}(T)$, then the resulting category $\text{Mod}_i(T)$ is a groupoid, i.e. a category in which every arrow has a two-sided inverse. Intuitively speaking, $\text{Mod}_i(T)$ is the category of models of T and their symmetries (i.e. automorphisms).

Since $\text{Mod}(T)$ doesn’t contain enough information to reconstruct T , a fortiori $\text{Mod}_i(T)$ doesn’t contain enough information to reconstruct T . To reiterate, a theory’s models and their automorphisms do not tell us everything about that theory! But now the insight of Awodey and Forssell was that if $\text{Mod}_i(T)$ is equipped with an appropriate topology, then the resulting topological groupoid G , could be the very G that appears in the representation theorem of Butz and Moerdijk. To be more precise, if C_T is the syntactic category of T , and if $Sh(C_T)$ is the topos of sheaves on C_T , then

$$Sh(C_T) \simeq B(G_T),$$

where G_T is the topological groupoid of models of T , and $B(G_T)$ is the Grothendieck topos of continuous actions of G_T . Furthermore, since the pretopos completion of C_T can be recovered as

²³Roughly, locales are topological spaces without a notion of a point, axiomatized instead with a primitive notion of a neighborhood and lattice operations on such neighborhoods (corresponding to unions and intersections). The advantage of locales is that they are amenable to a first-order axiomatization. The disadvantage is that they are strictly more general than topological spaces: every topological space is a locale but not every locale is a topological space. More precisely, it can be shown that the category of topological spaces is a coreflective subcategory of the category of locales. For this and more motivation on “pointless” topology see the introductory survey by Johnstone (1983).

the coherent objects in $Sh(C_T)$, it follows that the pretopos completion of C_T can be recovered from the topological groupoid G_T . In other words, T itself can be recovered from G_T up to Morita equivalence.

This result suggests that the syntactic and semantic categories of theories are dual to each other. On the one hand, we have **Pretop**, the category of (conceptually complete) syntactic categories. On the other hand, we have **TopGrpd**, the category of semantic categories, viz. topological groupoids. The “semantic functor” $\Theta : \mathbf{Pretop} \rightarrow \mathbf{TopGrpd}$ is defined by first taking $\mathbf{Pretop}(P, \mathcal{S})$, the category of set-valued models of P , then restricting to the isomorphisms between models, and finally equipping the resulting groupoid with the “logical” topology. The “syntactic functor” $\Gamma : \mathbf{TopGrpd} \rightarrow \mathbf{Pretop}$ is defined by taking a topological groupoid G to the topos $B(G)$ of continuous G -sets, and then extracting the pretopos of coherent objects in $B(G)$.²⁴

The two functors Θ and Γ are indeed adjoint to each other.

$$\begin{array}{ccc} & \Theta & \\ & \curvearrowright & \\ \mathbf{Pretop}^{op} & & \mathbf{TopGrpd} \\ & \curvearrowleft & \\ & \Gamma & \end{array}$$

However, once again, this adjunction is *not* an equivalence of categories; i.e. **TopGrpd** is not exactly

dual to **Pretop**. Thus, a natural question: can the semantic category \mathbf{Pretop}^{op} be characterized independently of the functor Θ ? i.e. can we provide an independent characterization of the category of (semantically presented) theories? Doing so would lead to a so-called “perfect duality” but such an independent characterization has so far proved elusive.²⁵

Why do mathematicians value duality results? One reason is that it enables them to transfer results and concepts from one category to its dual category. Thus, if we had a duality result

²⁴Technically, Awodey and Forssell work with the category **dCoh** rather than **Pretop**, and the results must be adjusted accordingly.

²⁵One reason for this – mysterious as it may sound – is certainly the fact that there seems to be nothing inherently “category-theoretic” about ultraproducts. As Makkai’s work proves and Los’ theorem has long made obvious, taking ultraproducts is a fundamental operation when it comes to elementary classes: elementary classes are exactly those classes closed under elementary equivalence and the taking of ultraproducts. Since every pretopos corresponds to an elementary class (more precisely: to the category of models of a coherent theory) one would imagine that any such characterization of **Pretop** would amount to a characterization of “closure under ultraproducts”. Absent any useful purely categorical description of ultraproducts (or even ultrafilters) this seems like a significant obstruction. Nevertheless the work of Leinster (2013) on ultrafilter monads as codensity monads might provide a way out, though this is still very far from being made precise.

for syntactic and semantic categories, then we could use information about theories as presented semantically in order to understanding theories as presented syntactically, and vice versa.

For example, suppose that the theory T' results from adding a new predicate symbol (but no new axioms) to the theory T . There is then an obvious translation of T into T' , namely the translation that takes each piece of non-logical vocabulary to itself. This translation corresponds to a functor $F : P_T \rightarrow P_{T'}$ that is faithful, full (since no new functional relations are created), but not essentially surjective (since no formula from the smaller language maps to the new predicate symbol). Recalling that F is an arrow in **Pretop**, there is a dual arrow $F^* : G_{T'} \rightarrow G_T$ in the semantic category **TopGrp**. It is natural to ask then: given that F has such and such features, what features does its dual arrow F^* have? In this case, the failure of essential surjectivity of F corresponds to the fact that F^* is not full, i.e. it forgets structure in the sense of (Weatherall, 2016).

Of course, the question can also be asked the other way around: given a functor $K : \mathbf{M} \rightarrow \mathbf{N}$ between categories of models, how do features of K correspond to features of its dual arrow K_* in the syntactic category? But here we pause, because in our opinion, the semantic category — whose objects are categories of models — has not yet been adequately characterized. First, not every topological groupoid is the groupoid of models of a coherent theory, i.e. **Pretop^{op}** has fewer objects than **TopGrpd**. Second, **Pretop^{op}** is a 2-category; hence, if **Pretop^{op}** is to be seen as living inside **TopGrpd**, then we must understand the latter as itself a 2-category. But what is the appropriate 2-categorical structure?²⁶

So, for anyone who wishes to develop the semantic view of theories, the following is a pressing question:

Given two categories of models \mathbf{M} and \mathbf{N} , what is a fruitful definition of an arrow $K : \mathbf{M} \rightarrow \mathbf{N}$? Furthermore, which arrows should be thought of as “reductions” of one theory to another, which as “equivalences” of theories, and which as other theoretical relations with which philosophers of science have been concerned?

Recalling what we said earlier: to choose the arrows of a category is to choose a notion of equivalence, and hence to choose a notion of identity of the category’s objects. (Here the objects of the semantic category are themselves categories, viz. *categories of models*.) Thus, until one proposes

²⁶Moerdijk (1988, 1990) defines a bicategory **LocGrpd** of localic groupoids, with bimodules as arrows, and shows that $G \mapsto B(G)$ is an equivalence of categories between **LocGrpd** and the bicategory of Grothendieck toposes.

a notion of arrows between semantic categories, then one lacks a clear notion of equivalence of theories, and hence of how a theory can be identified semantically. For philosophers of science, this issue demands immediate attention.

4 Conclusion

In this paper, we have assumed that it can be useful, for a philosophical understanding of science, to represent theories as mathematical objects. But what is a good, fruitful mathematical framework for understanding theories? We have surveyed a number of concepts and results from category theory, which we believe provide strong evidence that it should be the locus of attention for formal philosophers of science.

Firstly, we have shown that category theory provides the resources to get past philosophy of science's false dichotomy, viz. the dichotomy between syntactic versus semantic presentations of theories. On the one hand, a semantic presentation of a theory is nonetheless a presentation – written in a mathematical language. On the other hand, the syntactic category of a theory is a hybrid object, neither purely syntactic, nor purely semantic. Enough of the original syntax of theory can be reconstructed from its associated syntactic category. Moreover, having equivalent syntactic categories guarantees having equivalent categories of models. As such, the syntactic category of a theory unites the semantic and the syntactic approach to (first-order) theories, via the notion of Morita equivalence: as long as we care about theories up to (a suitable notion of) definitional equivalence and as long as we care about the classes of models of a theory up to its categorical structure, then the tension between syntactic and semantic presentations disappears. With this formal groundwork in place our most urgent task now is to carry as much of this lesson as possible over to the philosophy of science.

Secondly – with this false dichotomy set aside – philosophers of science are now set free from the illusion of a direct access to the thing in itself (via the class of models of a theory), and to study how different representations of the world can be related one to another. The category **Th** is the first formal approximation to how this kind of project might be carried out. The kind of notions about theoretical relations that will emerge will no doubt be interesting and we hope the philosophy of science lends a keen ear. Aside from this however, the very notion of studying scientific theories in their totality (i.e. as a structured whole) should lead to a philosophy of science more directly

attuned to the way theoretical physics is carried out today – where the interaction and interplay of (sometimes incompatible) theories is not seen merely as a temporary state of confusion as we approach some ultimate truth, but rather the necessary interactions that have to take place within any ecosystem before more advanced life can evolve out of it.

Acknowledgments: Thanks to Thomas Barrett for conversation and feedback.

References

- Andréka, H, JX Madarász, and I Németi (2001), “Defining new universes in many-sorted logic”, *Mathematical Institute of the Hungarian Academy of Sciences*, vol. 93.
- Awodey, Steve and Henrik Forssell (2013), “First-order logical duality”, *Annals of Pure and Applied Logic*, vol. 164, 3, pp. 319-348.
- Barr, Michael, John F. Kennison, and R. Raphael (2008), “Isbell duality”, *Theory and Applications of Categories*, vol. 20, 15, pp. 504-542.
- Barrett, Thomas and Hans Halvorson (2015a), “Glymour and Quine on theoretical equivalence”, unpublished manuscript, <http://philsci-archive.pitt.edu/11341/>.
- (2015b), “Morita equivalence”, <http://arxiv.org/abs/1506.04675>.
- (2015c), “Quine’s conjecture and a related confusion”, unpublished manuscript.
- Bénabou, Jean (1967), “Introduction to bicategories”, in *Reports of the Midwest Category Seminar*, pp. 1-77.
- Bickle, John (1998), *Psychoneural reduction: The new wave*, MIT Press.
- Borceux, Francis (1994), *Handbook of categorical algebra*, Cambridge University Press, vol. 1.
- Butz, Carsten and Ieke Moerdijk (1998), “Representing topoi by topological groupoids”, *Journal of Pure and Applied Algebra*, vol. 130, 3, pp. 223-235.
- Caramello, Olivia (2009), “Lattices of theories”, [arXiv:math.CT/0811.3547](https://arxiv.org/abs/math/0811.3547).
- (2010), “The unification of mathematics via topos theory”, [arXiv:1006.3930](https://arxiv.org/abs/1006.3930).
- (2012), “Universal models and definability”, *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 152, 2, pp. 279-302.
- Cassirer, Ernst (1953), *Philosophy of symbolic forms. Volume One: Language*, Yale University Press.

- Coecke, Bob and Aleks Kissinger (2016), “Quantum mechanics and ontology”, in *Categories for the working philosopher*, ed. by Elaine Landry, Oxford University Press.
- Corfield, David (2016), “??”, in *Categories for the working philosopher*, ed. by Elaine Landry, Oxford University Press.
- Curiel, Erik (2014), “Measure, topology and probabilistic reasoning in cosmology”, <http://philsci-archive.pitt.edu/11071/>.
- Fewster, Chris (2015), “Locally covariant quantum field theory and the problem of formulating the same physics in all spacetimes”, *Phil. Trans. A Roy. Soc.*, vol. 373, [arXiv:1502.04642](https://arxiv.org/abs/1502.04642).
- Fletcher, Samuel (2015), “Similarity, topology, and physical significance in relativity theory”, *The British Journal for Philosophy of Science*, [doi:10.1093/bjps/axu044](https://doi.org/10.1093/bjps/axu044).
- Forssell, Henrik (2013), “Subgroupoids and quotient theories”, *Theory and Applications of Categories*, vol. 28, 18, pp. 541-551.
- Friedman, Michael (1982), “The Scientific Image by Bas C. van Fraassen”, *The Journal of Philosophy*, vol. 79, 5, pp. 274-283.
- Gajda, Adam, Michał Krynicki, and Lesław Szčerba (1987), “A note on syntactical and semantical functions”, *Studia Logica*, vol. 46, 2, pp. 177-185.
- Halvorson, Hans (2012), “What scientific theories could not be”, *Philosophy of Science*, vol. 79, 2, pp. 183-206.
- (2015), “Scientific theories”, in *Oxford Handbook of the Philosophy of Science*, ed. by Paul Humphreys, Oxford University Press, <http://philsci-archive.pitt.edu/11347/>.
- Harnik, Victor (2011), “Model theory vs. categorical logic: Two approaches to pretopos completion (aka T^{eq})”, in *Models, logics, and higher-dimensional categories*, ed. by B. Hartt, T. Kucera, A. Pillay, and P.J. Scott, American Mathematical Society.
- Johnstone, Peter (1983), “The point of pointless topology”, *Bulletin of the American Mathematical Society*, vol. 8, 1, pp. 41-53.
- (2002), *Sketches of an elephant: A topos theory compendium*, Oxford University Press.
- Joyal, André and Myles Tierney (1984), “An extension of the Galois theory of Grothendieck”, *Memoires of the American Mathematical Society*, vol. 51, 309.
- Lack, Stephen (2010), “A 2-categories companion”, in *Towards higher categories*, ed. by John C. Baez and J. Peter May, Springer, pp. 105-191.

- Lal, Raymond and Nicholas J. Teh (2015), “Categorical generalization and physical structuralism”, *British Journal for the Philosophy of Science*, (forthcoming), <http://arxiv.org/abs/1404.3049>.
- Lambek, J. and P. J. Scott (1986), *Introduction to higher-order categorical logic*, Cambridge Studies in Advanced Mathematics 7, 7, Cambridge University Press.
- Leinster, Tom (2013), “Codensity and the ultrafilter monad”, *Theory and Applications of Categories*, vol. 28, pp. 332-370.
- MacLane, Saunders and Ieke Moerdijk (2012), *Sheaves in geometry and logic: A first introduction to topos theory*, Springer.
- Makkai, Michael (1987), “Stone duality for first order logic”, *Advances in Mathematics*, vol. 67, 2, pp. 97-170, doi:10.1016/0001-8708(87)90020-X.
- (1991), *Duality and definability in first order logic*, American Mathematical Soc., vol. 503.
- (1995), “On Gabbay’s proof of the Craig interpolation theorem for intuitionistic predicate logic”, *Notre Dame Journal of Formal Logic*, vol. 36, 3, pp. 364-381.
- Makkai, Michael and Gonzalo Reyes (1977), *First order categorical logic*, Springer-Verlag.
- Moerdijk, Ieke (1988), “The classifying topos of a continuous groupoid. I”, *Transactions of the American Mathematical Society*, vol. 310, 2, pp. 629-668.
- (1990), “The classifying topos of a continuous groupoid. II”, *Cahiers de Topologie et Géométrie Différentielle Catégoriques*, vol. 31, 2, pp. 137-168.
- Nagel, Ernest (1979), *The structure of science*, Hackett.
- Pitts, Andrew (1987), “Interpolation and conceptual completeness for Pretoposes via category theory”, in *Mathematical logic and theoretical computer science*, ed. by D. W. Kueker et al., M. Decker, New York.
- (1989), “Conceptual completeness for intuitionistic first-order logic: an application of categorical logic”, *Annals of Pure and Applied Logic*, vol. 41, pp. 33-81.
- Porst, H. E. and W. Tholen (1991), “Concrete dualities”, in *Category Theory at Work*, ed. by H. Heirlich and H. E. Porst, Heldermann Verlag Berlin, pp. 11-136.
- Reyes, Gonzalo (2009), “A derivation of Einstein’s vacuum field equations”, in *Models, logics, and higher-dimensional categories*, ed. by B. Hartt, T. Kucera, A. Pillay, and P.J. Scott, American Mathematical Society, pp. 245-262.

- Schreiber, Urs (2013), “Differential cohomology in a cohesive ∞ -topos”, <http://arxiv.org/abs/1310.7930>.
- Shulman, Mike (2012), “Exact completions and small sheaves”, *Theory and Applications of Categories*, vol. 27, 7, pp. 97-173.
- Teh, Nicholas J. and Dimitris Tsementzis (2015), “Theoretical equivalence in classical mechanics and its relationship to duality”, *Studies in the History and Philosophy of Modern Physics*, (forthcoming).
- Tsementzis, Dimitris (2015), “A syntactic characterization of Morita equivalence”, Manuscript, <http://arxiv.org/abs/1507.02302>.
- Van Benthem, Johan and David Pearce (1984), “A mathematical characterization of interpretation between theories”, *Studia Logica*, vol. 43, 3, pp. 295-303.
- Van Fraassen, Bas (1980), *The scientific image*, Oxford University Press.
- (2014), “One or two gentle remarks about Hans Halvorson’s critique of the semantic view”, *Philosophy of Science*, vol. 81, 2, pp. 276-283.
- Van Oosten, Jaap (2002), “Basic category theory”, <http://www.staff.science.uu.nl/~ooste110/syllabi/catsmoeder.pdf>.
- Weatherall, James Owen (2016), “Categories and the foundations of classical field theories”, in *Categories for the working philosopher*, ed. by Elaine Landry, Oxford University Press.