View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by PhilPapers

What Do Symmetries Tell Us About Structure?

Thomas William Barrett*

Abstract

Mathematicians, physicists, and philosophers of physics often look to the symmetries of an object for insight into the structure and constitution of the object. My aim in this paper is to explain why this practice is successful. In order to do so, I present a collection of results that are closely related to (and in a sense, generalizations of) Beth's and Svenonius' theorems.

1 Introduction

There is a famous idea about the relationship between the symmetries, or automorphisms, of a mathematical object and the structure of the object: An object's symmetries are often taken to provide us with significant information about its underlying structure. Hermann Weyl (1952, 144–5), for example, puts this idea as follows.

> A guiding principle in modern mathematics is this lesson: Whenever you have to do with a structure-endowed entity X, try to determine its group of automorphisms, the group of those element-wise transformations which leave all structural relations undisturbed. You can expect to gain a deep insight into the constitution of X in this way.

Mathematicians, physicists, and philosophers of physics often employ Weyl's guiding principle. But justification for it is rarely offered,[1] and one is therefore left to wonder exactly *why* the automorphisms of X provide us with insight into the constitution of X.

The aim of this paper is to substantiate the following simple idea: The symmetries of X provide us with a guide to which new structures are definable in terms of the "basic structure" of X. This idea yields a justification for

*Forthcoming in *Philosophy of Science*. I can be reached at thomaswbarrett@nyu.edu. I'm especially grateful to Hans Halvorson for many discussions about this material. Thanks to Neil Dewar, John Dougherty, Ben Feintzeig, Phillip Kremer, Alex Meehan, JB Manchak, and Jim Weatherall for comments and discussion on earlier versions of this paper. Thanks also to two anonymous referees for their helpful comments and feedback.

[1] A notable exception in philosophy of science is Dasgupta (2016). There are more exceptions in logic, where it is standard to recognize the kind of relationship between symmetry and definability that is the focus of this paper. For example, see the citations in footnote 3.

Weyl's guiding principle. It is natural to think of X as coming equipped with both its basic structure *and* those structures that the basic structure defines; these new structures intuitively "come for free given" or are "determined by" X's basic structure. The symmetries of X therefore provide us with insight into the constitution of X because they tell us which structures X is actually equipped with.

2 Symmetries and structure

We begin by discussing the standard method in which symmetries are used to examine an object. Philosophers of physics often trace this method back to the correspondence between Leibniz and Clarke on the nature of spacetime, and in particular, Leibniz's boost and shift arguments against Newtonian absolute space. In their modern gloss, Leibniz's arguments aim to show that particular pieces of structure that Newton is committed to — absolute position and absolute velocity — are not invariant under the symmetries of spacetime. Leibniz concludes from this that spacetime does not actually come equipped with those structure.

We now reason about symmetries and structure in an analogous manner: After determining the symmetries of a particular mathematical object X, one looks for the structures on X that are "invariant under" or "preserved by" all of these symmetries. Those structures that are found to be invariant under the symmetries of X are often deemed to be "determined by" or "constructed from" or "come for free given" the basic structure of X. On the other hand, those structures that are found to be *not* invariant under the symmetries of X are not accorded this same status.

One can grasp the basic idea behind this method by considering the following examples. The first two are examples of structures that are invariant under the symmetries of the underlying mathematical object, while the last two are examples of structures that are not.

Example 1. *The metric topology is invariant under the symmetries of a metric space.* Let (X, d) be a metric space, and consider the metric topology τ_d on X. Every automorphism of (X, d) preserves the topology τ_d, in the sense that it is a homeomorphism. ⌟

Example 2. *The Levi-Civita derivative operator is invariant under the symmetries of a manifold with metric.* Let (M, g_{ab}) be a smooth manifold with metric, and consider the Levi-Civita derivative operator ∇ associated with g_{ab}. Every automorphism of (M, g_{ab}) preserves the derivative operator ∇, in the sense that $f^*(\nabla_n \lambda^{a_1 \ldots a_r}_{b_1 \ldots b_s}) = \nabla_n f^*(\lambda^{a_1 \ldots a_r}_{b_1 \ldots b_s})$ for all smooth tensor fields $\lambda^{a_1 \ldots a_r}_{b_1 \ldots b_s}$. ⌟

Example 3. *An inner product is not invariant under the symmetries of a vector space.* Let V be a vector space, and consider an arbitrary inner product $\langle \cdot, \cdot \rangle$ on V. One can easily show that there is an automorphism of V that does not preserve $\langle \cdot, \cdot \rangle$. ⌟

Example 4. *The Galilean temporal metric is not invariant under the symmetries of Minkowski spacetime.* Let $(\mathbb{R}^4, \eta_{ab})$ be Minkowski spacetime, and consider the standard temporal metric $t_{ab} = (d_a x^1)(d_b x^1)$ of Galilean spacetime. There are automorphisms of $(\mathbb{R}^4, \eta_{ab})$ that do not preserve t_{ab} (Barrett, 2015b, Proposition 2). ⌟

There is a stark contrast between the first two examples and the last two. The metric topology τ_d and the Levi-Civita derivative operator ∇ are determined by the basic structure of their respective mathematical objects; there is a natural sense in which they come for free on (X, d) and (M, g_{ab}), respectively. In each case the basic structure of the mathematical object suffices to *define* the new piece of invariant structure. In Example 1, for instance, one uses the metric d to define the open balls in X, which one in turn uses to define the open sets of τ_d.

The structures on a mathematical object that are not invariant under the symmetries of the object, on the other hand, are not determined by the basic structure of the underlying mathematical object. In contrast to the invariant structures from Examples 1 and 2, the basic structure of the mathematical objects in Examples 3 and 4 does not suffice to define the new piece of structure. A vector space does not suffice to define a privileged inner product, and the Minkowski metric famously does not define a notion of absolute simultaneity on spacetime.

Examples like the four above suggest the following "conjecture" about the relationship between symmetry and structure:

Conjecture. *A piece of structure is invariant under the symmetries of a mathematical object if and only if it is definable from the basic structure of the object.*

If true, this conjecture would explain why Weyl's guiding principle is successful. It is natural to think of mathematical objects as coming equipped not only with their basic structure, but also with the structures that are definable in terms of their basic structure. Definable structures, like the metric topology and the Levi-Civita derivative operator, intuitively come for free given the basic structure on the object. So if the symmetries of an object tell us which structures on the object are definable, then they provide us with a guide to the structures that the object actually comes equipped with.

In order to consider whether this conjecture is true, we first need to clarify it. We do so by working in the framework of standard first-order logic. We will need some basic preliminaries.[2] A **signature** Σ is a set of predicate symbols, function symbols, and constant symbols. The Σ-terms, Σ-formulas, and Σ-sentences are recursively defined in the standard way. A **Σ-structure** A is a nonempty set in which the symbols of Σ have been interpreted. One recursively defines when a sequence of elements $a_1, \ldots, a_n \in A$ **satisfy** a Σ-formula $\phi(x_1, \ldots, x_n)$ in a Σ-structure A, written $A \vDash \phi[a_1, \ldots, a_n]$. We will use the notation ϕ^A to denote the set of tuples from the Σ-structure A that satisfy a Σ-formula ϕ. A

[2]The reader is encouraged to consult Hodges (2008) for further details.

Σ-sentence is a Σ-formula with no free variables. So if ϕ is a Σ-sentence, then $A \vDash \phi$ just in case the empty sequence satisfies ϕ in A. Two Σ-structures are **elementarily equivalent** if they satisfy precisely the same Σ-sentences. A **Σ-theory** T is a set of Σ-sentences. The sentences $\phi \in T$ are called the **axioms** of T. A Σ-structure M is a **model** of a Σ-theory T if $M \vDash \phi$ for all $\phi \in T$. Two Σ-theories are **logically equivalent** if they have the same class of models. A theory T **entails** a sentence ϕ, written $T \vDash \phi$, if $M \vDash \phi$ for every model M of T. If $\Sigma \subset \Sigma^+$ are signatures, we say that a Σ^+-theory T^+ is an **extension** of a Σ-theory T if $T \vDash \phi$ implies that $T^+ \vDash \phi$ for every Σ-sentence ϕ.

Given these preliminaries, there are two different approaches to making the conjecture precise that we will consider. The first approach begins with a Σ-structure A and asks whether the pieces of structure definable by A coincide with those that are invariant under the symmetries of A. Because of its focus on the Σ-structure A, which is often called a "model," we will call this the *model approach*. The second approach begins with a Σ-theory T and asks whether the pieces of structure definable by the theory T coincide with those that are invariant under the symmetries all of the models of T. Because of its focus on Σ-theories rather than Σ-structures, we will call this the *theory approach*. In what follows, our aim is to examine the senses in which the conjecture holds or fails. On both approaches, this depends on exactly how one explicates these notions of definability and invariance under symmetry. Sections 3 and 4 consider the two approaches in turn, before section 5 considers two payoffs that these results yield in philosophy of science.[3]

3 The model approach

The model approach begins with the following basic set-up.

- Let Σ be a signature. We think of the elements of Σ as the pieces of "basic structure" on the mathematical object under consideration.
- Let r be a symbol that is not contained in Σ. We think of r as the additional piece of structure that we are investigating. It may or may not be invariant under the symmetries of the mathematical object. Without loss of generality, we will assume that r is a unary predicate symbol.
- Let A be a $\Sigma \cup \{r\}$-structure. A is the mathematical object that we will be considering.

[3] Winnie (1986) contains all of the results from section 3. He does not present Theorems 1, 2, and 3 from section 4, however, and these results slightly extend his discussion. In addition, he does not directly address the two payoffs about structure and equivalence that we will discuss in section 5. In general, many of the results that follow are familiar to logicians (in addition to Winnie (1986), see Narens (2002), da Costa and Rodrigues (2007), and Korbmacher and Schiemer (2017) for example), but they unfortunately go unnoticed when symmetry is discussed in the context of physics. One of the aims of this paper, therefore, is to bring these logical results into contact with the current debate over the significance of symmetries in philosophy of physics.

This set-up relates to the above examples in a straightforward manner. In Example 1, "the language of metric spaces" plays the role of Σ, the metric topology plays the role of r, and the metric space (X, d) plays the role of A. In Example 3, "the language of vector spaces" plays the role of Σ, an inner product plays the role of r, and the vector space V plays the role of A.

Given this set-up, there is a particularly natural way to make precise what it means for the basic structure of A to define the additional piece of structure r. This is captured by the following condition.

($\mathbf{E}_A$) There is a Σ-formula ϕ such that $A \vDash \forall x(r(x) \leftrightarrow \phi(x))$.

When E_A holds we say that the Σ-structure A **explicitly defines** r in terms of Σ. This condition captures a clear sense in which the piece of structure r is "constructed from" the basic structures on A. The structure r is an "abbreviation" of some structure ϕ that A was already equipped with.

It is similarly easy to characterize when the piece of structure r is invariant under the symmetries of A.

($\mathbf{S}_A$) If $h : A|_\Sigma \to A|_\Sigma$ is an automorphism, then $h[r^A] = r^A$.

Here $A|_\Sigma$ is the Σ-structure obtained from A by forgetting the extension of the predicate r. An **automorphism** of a Σ-structure is a bijection from the object to itself that preserves the extensions of all of the predicates, functions, and constants in Σ. The automorphisms of $A|_\Sigma$ are just the maps from $A_|\Sigma$ to itself that preserve all of the basic structures in Σ. The condition S_A is therefore a straightforward way of saying that the symmetries of the basic structures of A preserve the structure r too.

The two conditions E_A and S_A provide a way to make the above conjecture precise. Indeed, if it is the case that E_A and S_A are equivalent, then the conjecture would be entirely substantiated. One direction is well-known.

Proposition 1. *If E_A, then S_A.*

Proof. Let ϕ be the Σ-formula (whose existence is guaranteed by E_A) that explicitly defines r, and let $h : A|_\Sigma \to A|_\Sigma$ be an automorphism. Since automorphisms preserve the extensions of all Σ-formulas, $h[\phi^A] = \phi^A$. E_A then guarantees that $\phi^A = r^A$, which immediately implies S_A. □

This proposition is most often used by appealing to its contrapositive, which provides a simple way of showing that a piece of structure is *not* definable. If we show that a piece of structure is not invariant under the symmetries of a mathematical object, then we can conclude that the object does not define that new structure from its basic structure. One can see this method in action by looking back to Example 3. Given an arbitrary inner product on a vector space V, one can easily show that there is an automorphism of V that does not preserve the inner product. The contrapositive of Proposition 1 (extrapolating beyond the simple case of first-order logic) implies that the inner product is *not* defined by the basic structure of the vector space V. So there is no natural inner product that is determined by the basic structure of a vector space.

Proposition 1 establishes a form of the "if" half of the conjecture, but it leaves open the "only if" half. And unfortunately, the following simple example demonstrates that this direction does not hold in general: *S_A does not imply E_A.*

Example 5. Let $\Sigma = \{c_0, c_1, c_2, c_3, \ldots\}$ be a signature containing a countable infinity of constant symbols. We want a $\Sigma \cup \{r\}$-structure A such that S_A holds, but E_A does not. In order to do this, we define the domain of A to be the countably infinite set $\{0, 1, 2, \ldots\}$ and we let $c_i^A = i$ for every i. Note that since there are uncountably many subsets of A, but only countably many Σ-formulas, there must be some subset of A that is not equal to ϕ^A for any Σ-formula ϕ. We let r^A be one such subset.

Now it is easy to verify that S_A holds of this $\Sigma \cup \{r\}$-structure A. Since every automorphism h of A satisfies $h(c_i^A) = c_i^A$, the only automorphism of A is the identity map. This immediately implies that S_A holds. But because of our choice of the subset r^A, there is no Σ-formula ϕ that satisfies $\phi^A = r^A$. Therefore E_A does not hold. ⌟

This example demonstrates a sense in which the "only if" half of the conjecture fails. In general, the symmetries of a mathematical object do not provide us with a complete guide to the structures that are definable from the basic structures of the object. It can be the case that a piece of structure is invariant under the symmetries of the object, but nonetheless fails to be definable. This happens in Example 5 simply because the only symmetry of A is the identity map. So *every* new piece of structure on A is invariant under its symmetries, regardless of whether or not it is definable from the basic structures of A.

There is nonetheless a way to establish a weaker form of the "only if" half of the conjecture. The symmetries of A alone do not encode which structures are definable on A, but we might allow ourselves to look to the symmetries of other objects that are closely related to A. The following result shows that this additional information completes the picture of definable structures on A (Winnie, 1986, Proposition 1.10).

Proposition 2. *E_A if and only if S_B holds for every $\Sigma \cup \{r\}$-structure B that is elementarily equivalent to A.*[4]

This proposition establishes a weaker form of the conjecture: A new piece of structure is definable on a mathematical object A if and only if the structure is invariant under the symmetries of every object that is elementarily equivalent to A.

Although the model approach does take a step towards substantiating the conjecture through Proposition 2, there are three reasons why it is worth investigating a different approach. First, a different approach might allow us to

[4] Winnie derives this result as a corollary to Svenonius' theorem, which we will discuss in the next section. Something slightly stronger is in fact the case: Proposition 2, Theorem 1, and Svenonius' theorem all have essentially the same content. Given any one of them we can easily prove either of the other two.

establish the conjecture in something closer to its full generality. Second, in philosophy of physics (and indeed in physics generally), one often wants to know what structure a *theory* as a whole posits. The theory approach that we will consider next is better suited to this task because of its focus on theories rather than on a specific Σ-structure. And lastly, the emphasis on theories yields corollaries that are relevant to two ongoing debates in philosophy of science about structure and equivalence. After examining the theory approach in detail, we will discuss these two debates.

4 The theory approach

The theory approach begins with a signature Σ and unary predicate symbol $r \notin \Sigma$, both of which are understood exactly as in the model approach. But rather than considering a $\Sigma \cup \{r\}$-structure A and examining the relationship between the structures it defines and the structures that are invariant under its automorphisms, the theory approach considers a $\Sigma \cup \{r\}$-theory T. One can think of T as picking out the "type of mathematical object" that will be under consideration. The approach then proceeds by examining the relationship between the structures that the theory T defines and those that are invariant under the symmetries of the models of T. This set-up relates to Examples 1–4 in much the same manner as the model approach did. In Example 4, for instance, "the language of Minkowski spacetime" plays the role of Σ, the Galilean temporal metric t_{ab} plays the role of r, and special relativity (or "the theory of Minkowski spacetime") plays the role of T.

In the model approach there was one natural way to make precise the idea that the structure r is definable in terms of those structures in Σ. In the theory approach, three ways suggest themselves. The first two are straightforward generalizations of E_A.[5]

(E1) There is a Σ-formula ϕ such that $T \vDash \forall x(r(x) \leftrightarrow \phi(x))$.

(E2) E_M holds for every model M of T.

When E1 holds, we will say that the theory T **globally explicitly defines** r in terms of Σ, and we call the sentence $\forall x(r(x) \leftrightarrow \phi(x))$ an **explicit definition** of r in terms of Σ. When E2 holds, on the other hand, we will say that the theory T **locally explicitly defines** r in terms of Σ.

Both E1 and E2 capture a sense in which the the new piece of structure r is constructed from the basic structures in Σ in every model of T. The difference between the two conditions is that E1 is requiring that the construction of r be uniform across models; the same Σ-formula ϕ must define r in every model of T. On the other hand, E2 allows for the possibility that *different* Σ-formulas define r in different models of T.

[5]Thanks to Phillip Kremer and an anonymous referee for pointing me towards E2.

It is easy to see that E1 implies E2. In order to show exactly how they are related, however, it will be useful to have a third explication of definability on the table.

(I1) For all models M and N of T, if $M|_\Sigma = N|_\Sigma$, then $r^M = r^N$.

When I1 holds, we say that T **implicitly defines** r in terms of Σ. This captures a sense in which r is determined by or "supervenes on" the basic structures in Σ. The condition simply says that whenever two models have the same basic structure, they must also agree on the structure r.

One might expect these different varieties of definability to be closely related. The following famous result establishes that this is indeed the case for global explicit definability and implicit definability (Hodges, 2008, Theorem 6.6.4).

Beth's theorem. *E1 if and only if I1.*

A simple example illustrates, however, that local explicit definability is a strictly weaker kind of definability: *E2 does not imply E1.*

Example 6. Let Σ be the empty signature and consider the $\Sigma \cup \{r\}$-theory T with the one axiom $\exists_{=1} x(x = x)$. This theory says that there is exactly one thing, but says nothing about whether it is r. The condition E2 holds of T. Indeed, if M is a model of T, then r^M is either the empty set or the entire (one element) domain of M. In the first case $M \vDash \forall x(r(x) \leftrightarrow x \neq x)$, while in the second case $M \vDash \forall x(r(x) \leftrightarrow x = x)$. E_M therefore holds of M, and since M was arbitrary E2 holds of T. But I1 trivially does not hold of T since there are models M and N with the same domain, but with $r^M \neq r^N$. Beth's theorem then implies that E1 also fails to hold of T. ⌟

We therefore have two different strengths of definability to consider. The stronger variety is captured by the conditions E1 and I1, while the weaker is captured by E2. With these explications of definability on the table, we once again turn to the conjecture from above. We would like to know what the relationship is between these three definability conditions and the invariance of r under symmetries. The following condition is a natural way to make precise the idea that r is invariant under symmetries of the basic structures in Σ.

(S1) For any model M of T, if $h : M|_\Sigma \to M|_\Sigma$ is an automorphism, then $h[r^M] = r^M$.

Note that S1 is simply requiring that S_M holds of every model M of T. The intuition behind the condition S_A therefore carries over to S1. Since automorphisms of a Σ-structure $M|_\Sigma$ are the maps from $M|_\Sigma$ to itself that preserve all of the basic structures in Σ, S1 is a straightforward way of saying that in every model of T the symmetries of the basic structures of M preserve the structure r too.

We begin by asking what the relationship is between S1 and the stronger variety of definability captured by conditions E1 and I1. One argues precisely as in Proposition 1 to demonstrate the following.

Proposition 3. *If E1, then S1.*

In conjunction with Beth's theorem, this result shows that I1 also implies S1. Proposition 3 gives us a form of the "if" half of the conjecture on the theory approach. As with Proposition 1, the most natural way to use Proposition 3 is by appealing to its contrapositive. If we can show that a piece of structure is not invariant under the symmetries of a particular model, then Proposition 3 licenses us to infer that the theory does not globally explicitly define that structure.[6]

As in the model approach, unfortunately, the "only if" half of the conjecture is not such a simple matter: *S1 does not imply E1.*

Example 6 (continued)**.** Recall the $\Sigma \cup \{r\}$-theory T defined in Example 6. We have seen that E1 does not hold of T. The condition S1, however, does hold. Indeed, if M is a model of T and $h : M|_\Sigma \to M|_\Sigma$ is an automorphism, then it must be that h is the identity map, since M only has one element. This immediately implies that $h[r^M] = r^M$. ⌟

The symmetries of models of a theory do not provide us with a complete guide to the structures that the theory globally defines on those models. Some structures that are invariant under the symmetries of every model nonetheless fail to be globally explicitly definable. Since local explicit definability is strictly weaker than global explicit definability, however, one naturally wonders whether the symmetries of models of a theory provide us with a complete guide to the structures that the theory *locally* defines. The following result establishes that this is the case.

Theorem 1. *E2 if and only if S1.*

Proof. Proposition 1 immediately establishes that E2 entails S1. Suppose that S1 holds and let M be a model of T. Let N be a $\Sigma \cup \{r\}$-structure that is elementarily equivalent to M. Since N must (trivially) be a model of T, S1 implies that S_N holds. So S_N holds for every $\Sigma \cup \{r\}$-structure N that is elementarily equivalent to M. Proposition 2 therefore implies that E_M holds. Since M was arbitrary, it must be that E2 holds of T. □

This result captures a form of the conjecture: The symmetries of models of a theory provide us with a complete guide to the structures that the theory *locally* explicitly defines. The way in which Theorem 1 substantiates the conjecture, however, leaves something to be desired. In particular, local explicit definability does not seem to be the kind of definability that is at play in standard cases like Examples 1 and 2. In those examples, the definition of the new structure does not vary between models; the metric topology, for instance, is defined in the same way regardless of the metric space that is under consideration.[7] For this

[6] In this respect the proposition is closely related to the "only if" half of Beth's theorem, which is sometimes called "Padoa's method."

[7] There is room for one to argue that a piece of structure that is merely locally explicitly definable does not "come for free" on the models of T. The models of the theory T from

reason, one might hope that more can be said about the relationship between *global* explicit definability and invariance under symmetry. And there is, in fact, a restricted class of theories whose symmetries do provide a perfect guide to global definability. We say that a Σ-theory T is **complete** if for every Σ-sentence ϕ, either $T \vDash \phi$ or $T \vDash \neg\phi$. When one restricts attention to complete theories, the converse of Proposition 3 holds (Hodges, 2008, Corollary 10.5.2).[8]

Svenonius' theorem. *If T is complete, then E1 if and only if S1.*

Svenonius' theorem captures yet another form of the conjecture. For complete theories, a piece of structure is globally explicitly definable if and only if it is invariant under the symmetries of every model. But this way of substantiating the conjecture again leaves something to be desired. Completeness is a strong condition to impose on a theory. Most theories are not complete, and one wonders what the relationship is between invariance under symmetry and global explicit definability for these more general theories.

Fortunately, there is a way to overcome this difficulty. So far, when asking what symmetries tell us about structure, we have only allowed ourselves to consider the automorphisms of a mathematical object. But automorphisms are just one particular kind of structure-preserving map — namely, the ones from an object to itself. This observation suggests a more general way to learn about the globally definable structures of a theory: Rather than only looking to the automorphisms of models of the theory, one can look to *all* of the structure-preserving maps between models of the theory. It is natural to wonder how much information this larger class of maps encodes about the structures that the theory defines.

The following two generalizations of the condition S1 are natural to consider.

(S2) For all models M and N of T, if $h : M|_\Sigma \to N|_\Sigma$ is an elementary embedding, then $h[r^M] = r^N$.

(S3) For all models M and N of T, if $h : M|_\Sigma \to N|_\Sigma$ is an isomorphism, then $h[r^M] = r^N$.

Two clarifications are in order about these conditions. First, an **elementary embedding** between Σ-structures A and B is a map $h : A \to B$ that satisfies

$$A \vDash \phi[a_1, \ldots, a_n] \text{ if and only if } B \vDash \phi[h(a_1), \ldots, h(a_n)]$$

for all Σ-formulas $\phi(x_1, \ldots, x_n)$ and elements $a_1, \ldots, a_n \in A$. And second, an **isomorphism** $h : A \to B$ between the Σ-structures A and B is a bijection

Example 6, for instance, do not seem to come equipped with the structure r in a particularly robust sense. This is an interesting issue, but I will set it aside for the purposes of this paper. It will suffice to say that, regardless of what status one wants to attribute to locally explicit definable structures, it is worth examining how much symmetries tell us about global explicit definability.

[8] Winnie (1986, Proposition 1.11) proves a conceptually similar result in the context of the model approach: For a restricted class of $\Sigma \cup \{r\}$-structures A (those that are "lucid"), E_A if and only if S_A.

that preserves the extensions of all predicates, functions, and constant symbols in Σ. Every automorphism is an isomorphism, and every isomorphism is an elementary embedding, but in general the converses do not hold.

The conditions S2 and S3 differ from S1 only in that they appeal to elementary embeddings and isomorphisms instead of automorphisms. These conditions provide two straightforward ways of saying that maps between models of T that preserve their basic structure also preserve the structure r. It turns out that S2 and S3 are both equivalent to E1 and I1. By themselves, the automorphisms of models of a theory do not provide us with all of the information about the globally definable structures on that object. But once we allow ourselves to look at these larger classes of maps — that is, elementary embeddings or isomorphisms — we have a perfect guide to globally definable structure.

Theorem 2. *E1 if and only if S2.*

Proof. Suppose first that E1 holds. Let M and N be models of T with $h : M|_\Sigma \to N|_\Sigma$ an elementary embedding. We immediately see that

$$h[r^M] = h[\phi^M] = \phi^N = r^N$$

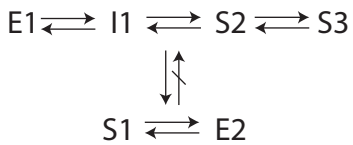
where ϕ is the Σ-formula (whose existence is guaranteed by E1) that explicitly defines r. The first and third equalities follow from E1, while the second equality holds since h is an elementary embedding. This implies S2.

Now suppose that S2 holds. Let M and N be models of T with $M|_\Sigma = N|_\Sigma$. The identity map $1 : M|_\Sigma \to N|_\Sigma$ is an elementary embedding, so by S2 it must be that $1[r^M] = r^N$. This immediately implies that $r^M = r^N$ and so $M = N$. We have therefore shown I1. Beth's theorem then implies E1. □

Theorem 3. *E1 if and only if S3.*

Proof. Suppose that E1 holds. Theorem 2 implies that S2 must hold, and since every isomorphism is an elementary embedding we immediately see that S3 holds too. On the other hand, if S3 holds, one establishes E1 by arguing exactly as in the "if" half of Theorem 2. □

The following figure summarizes the results from the theory approach.



We have two strengths of definability on the table. The weaker variety is represented by the condition E2, and the automorphisms of the models of a theory provide us with a perfect guide to this kind of definable structure. The stronger variety is represented by the conditions E1 and I1, and the class of structure-preserving maps between models of a theory provides us with a perfect guide to this kind.

5 Structure and equivalence

It is worth taking a moment to unravel how these results come to bear on the conjecture, which we restate here for convenience.

Conjecture. *A piece of structure is invariant under the symmetries of a mathematical object if and only if it is definable from the basic structure of the object.*

On both the model approach and the theory approach, the strongest form of the conjecture (and in particular, the "only if" half) fails. It is neither the case that E_A and S_A are equivalent, nor that E1 and S1 are equivalent. On the model approach, Proposition 2 captures a weaker form of the conjecture: If a piece of structure is invariant under the automorphisms both of X and all sufficiently closely related mathematical objects, then that structure must be definable.

On the theory approach, one can establish the following three weaker forms of the conjecture.

- If we weaken our notion of definability, then Theorem 1 substantiates the conjecture. The fact that E2 and S1 are equivalent shows that although symmetries do not provide us with a perfect guide to globally definable structure, they do provide us with a perfect guide to locally definable structure.
- If we restrict the class of theories that we are considering, then Svenonius' theorem substantiates the conjecture. The symmetries of models of *complete* theories tell us precisely which structures are globally explicit definable on the models.
- And lastly, if we allow ourselves to consider the entire class of structure-preserving maps between models of a theory, rather than merely the automorphisms of the models, then Theorems 2 and 3 substantiate the conjecture. This wider class of maps encodes all of the information about which structures are and are not globally explicitly definable on the models of the theory.

There is room for disagreement about which of these three comes closest to substantiating the original conjecture. I believe, however, that we have good reason to prefer the third. The first appeals to too weak a notion of definability, while the second is too restrictive about which theories the conjecture applies to. The third does not weaken our notion of definability, nor does it restrict the class of theories. Despite their differences, all three yield the same basic justification for Weyl's guiding principle about symmetry and structure: Symmetries provide insight into the structure and constitution of X because they tell us which structures are definable from X's basic structure.

It is common practice in philosophy of physics to appeal to a form Weyl's principle when examining the structure of our physical theories.[9] Two of the

[9]The principle is applied in philosophy of mathematics as well. Korbmacher and Schiemer

most well-known applications come from spacetime physics. First, symmetries are employed in debates between substantivalists and relationalists about the structure of spacetime. Relationalists will often argue that a particular piece of structure — like "absolute position" or "absolute velocity" — is not invariant under the symmetries of a spacetime theory. This type of argument can be understood as an appeal to the "if" half of the conjecture. If one can show that a piece of structure is not invariant under the symmetries of spacetime, then one is licensed to conclude that the structure is not definable in terms of the basic structure of spacetime. This in turn provides a strong sense in which spacetime simply does not come equipped with that structure.[10]

Second, there is a famous application of the "only if" half of the conjecture to the debate on the conventionality of simultaneity in special relativity. Philosophers of physics believed for many years that special relativity did not come equipped with a privileged notion of observer-relative simultaneity; the standard special relativistic notion of simultaneity was instead thought to be merely a convention. Malament (1977) was able to show, however, that the standard simultaneity relation on Minkowski spacetime is the only non-trivial equivalence relation that is invariant under the symmetries of special relativity. Malament explicitly appeals to the "only if" half of the conjecture to explain why his result is so powerful: It implies that the standard simultaneity relation is the only non-trivial equivalence relation that is definable in terms of the basic structure of Minkowski spacetime. In other words, Minkowski spacetime does not come equipped with any other candidate for a simultaneity relation.[11]

The conjecture also comes to bear directly on two issues in philosophy of science that have recently been under discussion: the question of how to compare amounts of structure between theories and the question of how to assess whether two theories are equivalent. We conclude by discussing these two topics in turn.

Structure

The history of classical spacetime theories is often viewed as a progression towards a "less structured" spacetime. Aristotelian spacetime posits more structure than Newtonian spacetime, which in turn posits more structure than Galilean spacetime.[12] In order to capture the relationship that these different spacetime theories bear to one another, one needs a precise method of comparing amounts of structure. Such a method would also be useful when diagnosing whether the models of a particular physical theory have "surplus structure"

(2017) discuss how the relationship between symmetries and definability comes to bear on structuralism.

[10] For discussion of spacetime symmetries, see Earman (1989), Dasgupta (2015, 2016), and the references therein. Weatherall (2017b) contains an argument about when the "extra facts" that the substantivalist demands are definable in a particular mathematical structure, connecting the issues discussed in this paper to some of the classic works on substantivalism, relationalism, and symmetry.

[11] See Winnie (1986) for a much more detailed discussion of Malament's result and of definability and symmetry in the broader context of spacetime and geometry.

[12] See Geroch (1978), Maudlin (2012), and Barrett (2015b) for discussion.

or when a theory is a "gauge theory."[13] There are two different methods of comparing amounts of structure that are currently on the table — the *automorphism approach* and the *category approach* — and the results above allow us to evaluate them against one another.

As its name suggests, the automorphism approach uses the automorphisms of mathematical objects to compare amounts of structure. Since the automorphisms of an object are the invertible structure-preserving maps from the object to itself, an object with "more automorphisms" intuitively must have "less structure" that these automorphisms are required to preserve. The amount of structure that an object has is (in some sense) inversely proportional to the size of the object's automorphism group. The automorphism approach is suggested by the discussions in Earman (1989) and North (2009), and Swanson and Halvorson (2012) and Barrett (2015a,b) make the approach precise by proposing the following criterion.

SYM*: X has more structure than Y if the automorphism group of X is a proper subset of the automorphism group of Y.

The above conjecture lends support to this kind of criterion. If an object has more automorphisms, then it is more difficult for a new piece of structure to be invariant under these automorphisms, so there will be fewer structures that are definable from the object's basic structure. And indeed, SYM* makes the intuitive verdicts when presented with many classic examples. A topological space has more structure than a bare set, an inner product space has more structure than a bare vector space, a manifold with metric has more structure than a bare manifold, and each of the classical spacetimes mentioned above has less structure than its predecessors (Barrett, 2015b).

But SYM* has a serious shortcoming, stemming from the fact that S_A does not entail E_A. The automorphisms of an object do not provide a perfect guide to definable structures on that object. Example 5 makes this shortcoming precise. Consider the $\Sigma \cup \{r\}$-structure A from Example 5 and its reduct $A|_\Sigma$ to the signature Σ. It is perfectly natural to think that the mathematical object A has more structure than $A|_\Sigma$. Indeed, $A|_\Sigma$ is obtained by *forgetting the structure* r from the object A, and furthermore, r does not come for free on $A|_\Sigma$ since E_A does not hold. So A has a piece of structure r that $A|_\Sigma$ lacks. But according to SYM* it is not the case that A has more structure than $A|_\Sigma$; the two objects have the same trivial automorphism group. This is an undesirable verdict. One therefore hopes that SYM*, and the automorphism approach in general, can be improved upon.

Theorems 2 and 3 suggest one potential improvement: We can obtain a better guide to the amount of structure that an object has by looking to the class of *all* structure-preserving maps between objects rather than merely the automorphisms. In fact, a method of comparing amounts of structure that employs exactly this idea is already on the table. The category approach to comparing amounts of structure was originally proposed by Baez et al. (2006),

[13] See Weatherall (2016b) and Nguyen et al. (2017).

and has recently been been employed in philosophy of physics by Weatherall (2016b, 2017a), Nguyen et al. (2017), and Feintzeig (2017). As its name suggests, the idea behind this approach is that one can compare amounts of structure between mathematical objects by looking to the *categories* in which the objects reside.

In order to explain this method of comparing amounts of structure, we need the following simple category-theoretic machinery.[14] A first-order theory T has a category of models. A **category** C is a collection of objects with arrows between the objects that satisfy some basic properties. We will use the notation $\mathrm{Mod}(T)$ to denote the **category of models** of T. An object in $\mathrm{Mod}(T)$ is a model M of T, and an arrow $f : M \to N$ between objects in $\mathrm{Mod}(T)$ is an elementary embedding $f : M \to N$ between the models M and N. A functor $F : C \to D$ between categories C and D is a structure-preserving map between categories. When T^+ is an extension of a Σ-theory T, we can define the functor $\Pi : \mathrm{Mod}(T^+) \to \mathrm{Mod}(T)$ by

$$\Pi(M) = M|_\Sigma \qquad \Pi(h) = h$$

for every model M of T^+ and elementary embedding h between models of T^+. We say that a functor $F : C \to D$ is **full** if for all objects c_1, c_2 in C and arrows $g : Fc_1 \to Fc_2$ in D there exists an arrow $f : c_1 \to c_2$ in C with $Ff = g$. F is **faithful** if for all objects c_1, c_2 in C and arrows $f, g : c_1 \to c_2$, $Ff = Fg$ implies that $f = g$. And F is **essentially surjective** if for every object d in D there is an object c in C such that Fc is isomorphic to d. A functor that is full, faithful, and essentially surjective is called an **equivalence** of categories.

Baez et al. (2006) classify functors between categories based on "what they forget." Most importantly for our purposes, when a functor $F : C \to D$ is not full it is said to **forget structure**. The existence of a functor $F : C \to D$ that forgets structure captures a sense in which (relative to the comparison generated by F) objects of D have less structure than objects of C. One can see the idea behind this method by considering the following example. It is standard to recognize a sense in which topological spaces have more structure than sets, and the category approach allows one to recover this sense.

Example 7. Consider the categories Set and Top. The objects of Set are sets and the arrows are functions between sets. The objects of Top are topological spaces and the arrows are continuous functions. One particularly natural functor $U : \mathrm{Top} \to \mathrm{Set}$ is defined by

$$U : (X, \tau) \longmapsto X \qquad U : f \longmapsto f$$

for all topological spaces (X, τ) and continuous functions f. One can easily verify that U is a functor. It converts a topological space into a set by "forgetting" about the topology. Since there are functions between some topological spaces that are not continuous, U trivially is not full and therefore forgets structure. ⌟

[14] The reader is encouraged to consult Mac Lane (1971) or Borceux (1994) for further details. We take for granted the definitions of a category and of a functor.

The general motivation behind the category approach is essentially the same as that behind automorphism approach. Since the functor $U : \text{Top} \to \text{Set}$ is not full, this provides a sense in which there are "more arrows" (relative to the comparison given by U) between objects in the category Set than there are between objects in the category Top. The arrows in these categories are structure-preserving maps between the objects. Therefore, since there are "more structure-preserving maps" between the objects of Set than there are between the objects of Top, the former must have less structure that these maps are required to preserve.

Theorem 2 yields a corollary that takes a first step towards justifying the category approach.

Corollary 1. *Let T^+ be a $\Sigma \cup \{r\}$-theory that is an extension of the Σ-theory T. The functor $\Pi : Mod(T^+) \to Mod(T)$ forgets structure if and only if E1 does not hold of T^+.*

Proof. It is easy to verify that Π is full if and only if S2 holds of T^+. Theorem 2 then immediately implies the corollary. □

If one extrapolates beyond the case of first-order theories, this corollary tells us that a functor from C to D forgets structure if and only if the objects in C have structure that is not definable from the structure of the objects in D.[15]

The fact that S_A does not imply E_A generates uncomfortable counterexamples to the automorphism approach and SYM* in particular. This is what saw in Example 5. The example does not, however, generate a problem for the category approach. Let T^+ be the $\Sigma \cup \{r\}$-theory that has as axioms every $\Sigma \cup \{r\}$-sentence ϕ such that $A \vDash \phi$, and let T be the Σ-theory that has as axioms every Σ-sentence ψ such that $A|_\Sigma \vDash \psi$. Since E_A does not hold, it is easy to see that E1 does not hold of T^+. Corollary 1 therefore implies that Π forgets structure, capturing a sense in which models of T^+ (like A) have more structure than models of T (like $A|_\Sigma$). The category approach therefore makes intuitive verdicts in some cases where the automorphism approach stumbles.

And more generally, the equivalence of E1 with S2 and S3 suggests that the category approach is a conceptual improvement upon the automorphism approach. The automorphism group of a mathematical object does not provide us with a perfect guide to the amount of definable structure that an object has. But Theorems 2 and 3, along with Corollary 1, suggest that the *category* in which the object resides does provide us with such a guide.

[15] There are two reasons why this extrapolation is not yet completely justified. First, Corollary 1 only concerns the projection functor Π, which is more well-behaved than an arbitrary functor. And second, Corollary 1 only concerns the first-order case. In higher-order logics — which are discussed by Winnie (1986) and da Costa and Rodrigues (2007) — Beth's theorem does not hold. Since it plays a crucial role in our proof of Corollary 2, this proof will not easily generalize to the higher-order case. More work is therefore required before the category approach is completely justified.

Equivalence

The results also come to bear on the recent debate about theoretical equivalence.[16] We would like to know the conditions under which two theories should be considered equivalent. Many criteria for equivalence have been proposed, but two will be of particular interest to us here: definitional equivalence and categorical equivalence.

Definitional equivalence was first introduced into philosophy of science by Glymour (1971, 1977, 1980). The concept is simple given our discussion of definability above. Let $\Sigma \subset \Sigma^+$ be signatures. A **definitional extension** of a Σ-theory S to the signature Σ^+ is a Σ^+-theory that is logically equivalent to the theory

$$S^+ = S \cup \{\delta_s : s \in \Sigma^+ - \Sigma\},$$

where for each symbol $s \in \Sigma^+ - \Sigma$, the sentence δ_s is an explicit definition of s in terms of Σ. Two theories are **definitionally equivalent** if they have a common definitional extension.[17]

It has also been suggested by Weatherall (2016a) and Halvorson (2016), among others, that category theory provides us with a standard for equivalence between theories: Two theories T_1 and T_2 are **categorically equivalent** if their categories of models $\mathrm{Mod}(T_1)$ and $\mathrm{Mod}(T_2)$ are "structurally identical," i.e. if there is a functor between them that is an equivalence of categories. Both definitional equivalence and categorical equivalence are supposed to capture senses in which two theories might be considered "intertranslatable."

Unfortunately, both of these criteria suffer from some shortcomings. Weatherall (2016a) and Barrett and Halvorson (2016b) have argued that definitional equivalence is too strict a standard of equivalence. It judges theories to be inequivalent that we have good reason to consider equivalent. And on the other hand, categorical equivalence is too liberal. Barrett and Halvorson (2016b, Theorem 5.2) provide the following example of categorically equivalent theories T_1 and T_2 that we nonetheless have good reason to consider inequivalent.

Example 8. Consider the two signatures $\Sigma_1 = \{p_0, p_1, p_2, p_3 \ldots\}$ and $\Sigma_2 = \{q_0, q_1, q_2, \ldots\}$, each of which have a countable infinity of unary predicate symbols. We define the Σ_1-theory T_1 and the Σ_2-theory T_2 as follows.

$$T_1 = \{\exists_{=1} x(x = x)\}$$
$$T_2 = \{\exists_{=1} y(y = y), \forall y(q_0(y) \to q_1(y)), \forall y(q_0(y) \to q_2(y)), \ldots\}$$

One can prove that T_1 and T_2 are categorically equivalent, but they are not definitionally equivalent (Barrett and Halvorson, 2016b, Theorem 5.2). ⌟

[16] North (2009), Knox (2014), Curiel (2014), Barrett (2015a, 2017), Hudetz (2015), Rosenstock et al. (2015), Weatherall (2016a, 2017a), and Rosenstock and Weatherall (2016) discuss whether particular physical theories are equivalent. For discussion of different criteria for equivalence see Barrett and Halvorson (2016b), Hudetz (2017), and the references therein.

[17] Definitional equivalence has received attention from logicians for many years. For example, see de Bouvére (1965), Kanger (1968), Pinter (1978), Pelletier and Urquhart (2003), Andréka et al. (2005), Friedman and Visser (2014), and Barrett and Halvorson (2016a,b, 2017a,b), and the references therein.

Although they are categorically equivalent, there is a strong sense in which these two theories are inequivalent. T_1 says that there is one thing, while T_2 says that there is one thing, and in addition, that there is a "special predicate" q_0 with the following property: If q_0 holds of the one thing, then all of the predicates q_i must hold of the one thing too. Accordingly to T_1, there is no such special predicate. These two theories are not "saying the same thing," so it is a mark against categorical equivalence that it judges them to be equivalent.

Definitional equivalence judges them to be inequivalent. T_1 and T_2 fail to be definitionally equivalent because T_2 cannot define the special predicate q_0. This points to a certain intuitive and desirable feature that definitional equivalence has but categorical equivalence lacks: If two theories are definitionally equivalent, then one can "build" or "construct" the models of the one theory from the models of the other, and vice versa. Example 8 shows that categorical equivalence does not have this feature; there are theories that are categorically equivalent despite the fact that one cannot construct models of T_2 from models of T_1. This discussion demonstrates the following:

> It is not the case that an equivalence of categories between $\mathrm{Mod}(T_1)$ and $\mathrm{Mod}(T_2)$ implies that the structures of T_1 are definable in terms of the structures of T_2.

This result is a definite mark against categorical equivalence as a general standard for equivalence of theories. It does not capture as robust a notion of intertranslatability as one might hope. Accordingly, one would like to strengthen categorical equivalence so that it might better capture facts about definability.[18]

The results here suggest that this is indeed possible. In particular, we have the following simple corollary.

Corollary 2. *Let T^+ be a $\Sigma \cup \{r\}$-theory that is an extension of the Σ-theory T. The functor $\Pi : Mod(T^+) \to Mod(T)$ is an equivalence if and only if T^+ is a definitional extension of T.*

Proof. The proof of the "if" direction is familiar; it follows from Theorem 5.1 of Barrett and Halvorson (2016b). Assume then that Π is an equivalence. Since Π is full, Corollary 1 implies that E1 holds of T^+, so

$$T^+ \vDash \forall x(\phi(x) \leftrightarrow r(x))$$

for some Σ-formula ϕ. Now using the fact that Π is essentially surjective, one easily verifies that $T \cup \{\forall x(\phi(x) \leftrightarrow r(x))\}$ is logically equivalent to T^+. □

[18] It is an open question, however, just *how many* theories there are like the pair from Example 8, or in other words, *how much weaker* categorical equivalence is than definitional equivalence. Looking for a strengthened variety of categorical equivalence will necessarily yield progress on this question too. Hudetz (2017) makes progress on both counts. The more general version of this question is how much stronger "Morita equivalence" — the many-sorted analogue of definitional equivalence — is than categorical equivalence. We will set aside many-sorted concerns here, but the reader is invited to consult Barrett and Halvorson (2016b) for details.

Example 8 showed that the existence of an arbitrary equivalence between the categories of models of two theories does not guarantee that the two theories can define one another's structures. But if Π is an equivalence, then Corollary 2 implies that the two theories can do precisely this. It is natural to wonder whether there is some special property $\mathfrak{P}$ of the functor Π that allows it to encode more about definable structure than an arbitrary functor does. This suggests a family of conjectures of the following form:

> If there is a functor F that (i) is an equivalence of categories between $\mathrm{Mod}(T_1)$ and $\mathrm{Mod}(T_2)$ and (ii) has property $\mathfrak{P}$, then T_1 and T_2 are definitionally equivalent.[19]

Results of this form would improve upon categorical equivalence as a general standard of equivalence between theories. Steps in exactly this direction are currently being taken by Hudetz (2017).

6 Conclusion

The results here about definability and symmetry begin to provide justification for the use of category theoretic tools when examining the relationships between theories. These tools seem to be particularly well-suited for capturing when models of one theory have less structure than models of another theory, and when two theories are equivalent. More generally, these results provide support for one of the primary motivations behind category theory.[20] The idea at the heart of category theory is simple: Mathematical objects can be thought of, not in terms of their "internal structure," but rather in terms of the relations that they bear to other objects. For example, from the category theoretic perspective one sees a group not as a set with a binary operation, but instead as an object in a particular network of arrows. This viewpoint has proven useful over the course of the last sixty years, yielding applications in mathematics, computer science, and physics. The extent to which the perspective is justified, however, depends on precisely how much information about mathematical objects is encoded by the arrows — that is, the structure-preserving maps — between the objects. Theorems 2 and 3 take a step towards justifying this perspective, and in doing so, suggest a fruitful generalization of Weyl's guiding principle: *One can gain insight into the constitution of a mathematical object by looking to the class of all structure-preserving maps between objects of the same kind.*

[19] As in the previous footnote, the more general conjecture replaces definitional equivalence with Morita equivalence.

[20] Winnie (1986) also discusses the relationship between definability, invariance, and category theory, but focuses more on the concept of a natural transformation than on categorical equivalence.

References

Andréka, H., Madarász, J. X., and Németi, I. (2005). Mutual definability does not imply definitional equivalence, a simple example. *Mathematical Logic Quarterly*, 51(6):591–597.

Baez, J., Bartels, T., Dolan, J., and Corfield, D. (2006). Property, structure and stuff. Available at http://math.ucr.edu/home/baez/qg-spring2004/discussion.html.

Barrett, T. W. (2015a). On the structure of classical mechanics. *The British Journal for the Philosophy of Science*, 66(4):801–828.

Barrett, T. W. (2015b). Spacetime structure. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 51:37–43.

Barrett, T. W. (2017). Equivalent and inequivalent formulations of classical mechanics. *Forthcoming in the British Journal for the Philosophy of Science*.

Barrett, T. W. and Halvorson, H. (2016a). Glymour and Quine on theoretical equivalence. *Journal of Philosophical Logic*, 45(5):467–483.

Barrett, T. W. and Halvorson, H. (2016b). Morita equivalence. *The Review of Symbolic Logic*, 9(3):556–582.

Barrett, T. W. and Halvorson, H. (2017a). From geometry to conceptual relativity. *Erkenntnis*, 82(5):1043–1063.

Barrett, T. W. and Halvorson, H. (2017b). Quine's conjecture on many-sorted logic. *Synthese*, 194(9):3563–3582.

Borceux, F. (1994). *Handbook of Categorical Algebra*, volume 1. Cambridge University Press.

Curiel, E. (2014). Classical mechanics is Lagrangian; it is not Hamiltonian. *The British Journal for the Philosophy of Science*, 65(2):269–321.

da Costa, N. C. A. and Rodrigues, A. A. M. (2007). Definability and invariance. *Studia Logica*, 86(1):1–30.

Dasgupta, S. (2015). Substantivalism vs relationalism about space in classical physics. *Philosophy Compass*, 10(9):601–624.

Dasgupta, S. (2016). Symmetry as an epistemic notion (twice over). *The British Journal for the Philosophy of Science*, 67(3):837–878.

de Bouvére, K. L. (1965). Synonymous theories. In *Symposium on the Theory of Models*, pages 402–406. North-Holland Publishing Company.

Earman, J. (1989). *World Enough and Spacetime: Absolute versus Relational Theories of Space and Time*. MIT.

Feintzeig, B. H. (2017). Deduction and definability in infinite statistical systems. *Forthcoming in Synthese*.

Friedman, H. M. and Visser, A. (2014). When bi-interpretability implies synonymy. *Logic Group Preprint Series*, 320:1–19.

Geroch, R. (1978). *General Relativity from A to B*. Chicago University Press.

Glymour, C. (1971). Theoretical realism and theoretical equivalence. In *PSA 1970*, pages 275–288. Springer.

Glymour, C. (1977). The epistemology of geometry. *Noûs*, 11:227–251.

Glymour, C. (1980). *Theory and Evidence*. Princeton University Press.

Halvorson, H. (2016). Scientific theories. In Humphreys, P., editor, *The Oxford Handbook of Philosophy of Science*, pages 585–608. Oxford University Press.

Hodges, W. (2008). *Model Theory*. Cambridge University Press.

Hudetz, L. (2015). Linear structures, causal sets and topology. *Studies in History and Philosophy of Modern Physics*, pages 294–308.

Hudetz, L. (2017). Definable categorical equivalence: Towards an adequate criterion of theoretical intertranslatability. *Forthcoming in Philosophy of Science*.

Kanger, S. (1968). Equivalent theories. *Theoria*, 34(1):1–6.

Knox, E. (2014). Newtonian spacetime structure in light of the equivalence principle. *The British Journal for the Philosophy of Science*, 65(4):863–880.

Korbmacher, J. and Schiemer, G. (2017). What are structural properties? *Forthcoming in Philosophia Mathematica*.

Mac Lane, S. (1971). *Categories for the working mathematician*. Springer.

Malament, D. B. (1977). Causal theories of time and the conventionality of simultaneity. *Noûs*, pages 293–300.

Maudlin, T. (2012). *Philosophy of Physics: Space and Time*. Princeton University Press.

Narens, L. (2002). *Theories of Meaningfulness*. Lawrence Erlbaum Associates.

Nguyen, J., Teh, N. J., and Wells, L. (2017). Why surplus structure is not superfluous. *Forthcoming in the British Journal for the Philosophy of Science*.

North, J. (2009). The 'structure' of physics: A case study. *The Journal of Philosophy*, 106:57–88.

Pelletier, F. J. and Urquhart, A. (2003). Synonymous logics. *Journal of Philosophical Logic*, 32(3):259–285.

Pinter, C. C. (1978). Properties preserved under definitional equivalence and interpretations. *Mathematical Logic Quarterly*, 24(31-36):481–488.

Rosenstock, S., Barrett, T. W., and Weatherall, J. O. (2015). On Einstein algebras and relativistic spacetimes. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 52:309–316.

Rosenstock, S. and Weatherall, J. O. (2016). A categorical equivalence between generalized holonomy maps on a connected manifold and principal connections on bundles over that manifold. *Journal of Mathematical Physics*, 57(10). arXiv:1504.02401 [math-ph].

Swanson, N. and Halvorson, H. (2012). On North's 'The structure of physics'. *Manuscript*.

Weatherall, J. O. (2016a). Are Newtonian gravitation and geometrized Newtonian gravitation theoretically equivalent? *Erkenntnis*, 81(5):1073–1091.

Weatherall, J. O. (2016b). Understanding gauge. *Philosophy of Science*, 83(5):1039–1049.

Weatherall, J. O. (2017a). Category theory and the foundations of classical field theories. In Landry, E., editor, *Forthcoming in Categories for the Working Philosopher*. Oxford University Press.

Weatherall, J. O. (2017b). Regarding the 'hole argument'. *Forthcoming in the British Journal for the Philosophy of Science*.

Weyl, H. (1952). *Symmetry*. Princeton University Press.

Winnie, J. (1986). Invariants and objectivity: A theory with applications to relativity and geometry. In Colodny, R. G., editor, *From Quarks to Quasars*, pages 71–180. Pittsburgh: Pittsburgh University Press.