DOI: 10.1111/nous.12358

ARTICLE

NOÛS

How to count structure

Thomas William Barrett

Department of Philosophy, University of California, Santa Barbara

Correspondence
Thomas William Barrett, Department of Philosophy, University of California, Santa Barbara.
Email: thomaswbarrett@ucsb.edu

Abstract

There is sometimes a sense in which one theory posits 'less structure' than another. Philosophers of science have recently appealed to this idea both in the debate about equivalence of theories and in discussions about structural parsimony. But there are a number of different proposals currently on the table for how to compare the 'amount of structure' that different theories posit. The aim of this paper is to compare these proposals against one another and evaluate them on their own merits.

1 | INTRODUCTION

The history of classical spacetime theories is often presented as a progression towards a 'less structured' spacetime.[1] Indeed, the story is often told as follows.

> We began long ago with Aristotelian spacetime. Aristotelian spacetime singles out a preferred location as the *center of the universe*. When we moved to Newtonian spacetime we did away with this structure. Newtonian spacetime does not single out a preferred location, but it does single out a preferred inertial frame as *absolute rest*. Finally, we moved to Galilean spacetime and again did away with structure. Galilean spacetime does not even single out a preferred rest frame.

It is standard to draw the following conclusion from this story. It is, in fact, implicit in the way the story is usually told.

Conclusion 1. Each of these classical spacetime theories *posits less structure*, or *ascribes less structure to the world*, than its predecessors. Galilean spacetime, for example, is obtained by 'taking something away' — namely, the concept of absolute rest — from Newtonian spacetime.

[Correction added on 1 December 2020, after first online publication: the line diagram is added to the appendix section.]

The aim of this paper is to better understand claims like Conclusion 1. In particular, the aim is to understand what this relationship of 'positing less structure than' amounts to.

While Conclusion 1 is itself intuitive, there are more difficult cases outside the realm of classical spacetime theories that one might want to consider. For example, North (2009) has recently argued that the Hamiltonian formulation of classical mechanics posits less structure than the Lagrangian formulation. And Rosenstock, Barrett, and Weatherall (2015) have argued that, contrary to what Earman (1986, 1989) claims, the standard geometric formulation of general relativity does not posit more structure than its algebraic formulation, which appeals to the mathematical apparatus of Einstein algebras. Since there are these more difficult cases of structural comparison, we want to have some general criteria that we can use to tell whether one theory posits less structure than another.

There are two main reasons why this structural relationship is philosophically important. The first has to do with another inter-theoretic relation that philosophers have recently been examining: equivalence of theories. It is natural to draw the following corollary from Conclusion 1.

Conclusion 2. Each of these classical spacetime theories is *not equivalent* to its predecessors. Aristotelian and Newtonian spacetime, for example, disagree about whether or not there is a center of the universe.

Equivalent theories are supposed to be mere 'notational variants' of one another; they say the same thing about the world, but say it in different ways. They might, for example, be formulated using different languages or different mathematics, but all of the differences between them are taken to be inconsequential. The two most cited examples of equivalent theories in physics are the Heisenberg and Schrödinger formulations of quantum mechanics and the Hamiltonian and Lagrangian formulations of classical mechanics. But the general question of when we should consider two theories to be equivalent — and indeed, whether or not these purported examples are actual cases of equivalence — has recently been the subject of significant debate.[2]

The story of classical spacetime theories — and in particular the move from Conclusion 1 to Conclusion 2 — suggests a particular way to approach questions of equivalence. Since the Aristotelian, Newtonian, and Galilean theories of spacetime all ascribe different amounts of structure to spacetime, they all say different things about the world, and therefore must be inequivalent theories. They disagree about the amount of structure that the world has. In general, this kind of reasoning about amounts of structure should carry over when we consider whether other theories are equivalent: If two theories posit different amounts of structure, then they must be inequivalent. This method of answering questions of equivalence has, in fact, already been employed. North (2009) infers from her claim that Hamiltonian mechanics posits less structure than Lagrangian mechanics that the two theories must actually be *in*equivalent, dissenting from the standard view. This is one reason why it is important for us to better understand claims like Conclusion 1. We would thereby gain a tool that we can use to judge whether two theories are equivalent.

There is a second reason why it is important to understand claims like Conclusion 1. The story about the progression of classical spacetime theories has motivated some metaphysicians, philosophers of physics, and physicists to adopt a version of the following methodological principle.

Structural parsimony. All other things equal, we should prefer theories that posit less structure.

North (2009, p. 64), for example, puts this idea as follows:

> This is a principle informed by Ockham's razor; though it is not just that, other things being equal, it is best to go with the ontologically minimal theory. It is not that, other things being equal, we should go with the fewest entities, but that we should go with the least structure.

Sider (2013, p. 240) argues that "'structurally simpler' theories are more likely to be true", which would certainly give us good reason to prefer them. Earman (1989, p. 46) argues that we should avoid theories that use "more space-time structure than is needed to support the laws". Friedman (1983, p. 112) argues the same. And even the mathematical physicist Geroch (1978, p. 52) comes close to endorsing the structural parsimony principle. After discussing the transition from Newtonian to Galilean spacetime, he writes:

> Although the evidence on this is perhaps a bit scanty, it seems to be the case that physics, at least in its fundamental aspects, always moves in this one direction. It may not be a bad rule of thumb to judge a new set of ideas in physics by the criterion of how many of the notions and relations that one feels to be necessary one is forced to give up.

Even though the structural parsimony principle is often endorsed, it stands in need of clarification. One needs to clarify exactly what the other things are, what it might mean for them to be equal, and exactly what kind of preference the principle licenses. The most essential clarification that must be made, however, is the one that we intend to pursue in this paper. In order to even say which pairs of theories the structural parsimony principle is applicable to, one needs to understand the conditions under which one theory posits less structure than another. In other words, one needs to understand how to count, or compare, amounts of structure. A number of different proposals for how to compare amounts of structure have recently been put forward.[3] But there has so far been no systematic evaluation of them. The aim of this paper is to provide such an evaluation.

The proposals naturally divide into two broad approaches: the automorphism approach and the category approach. The former tries to answer the question of how to compare amounts of structure by looking to the automorphisms, or symmetries, of the objects under consideration, while the latter tries to answer the question by looking to the categories in which the objects reside. Both have recently been employed by philosophers of physics. After evaluating these two approaches in detail, I will conclude by returning to these more general issues of structural parsimony and equivalence. Our discussion will yield two modest payoffs. The first concerns the conditions under which the structural parsimony principle is applicable; the second concerns the conditions under which we are licensed to infer the inequivalence of two theories from apparent structural differences between them. In both of these cases, the conditions are harder to satisfy than one might have initially thought.

2 | THE AUTOMORPHISM APPROACH

It is useful to begin with some simple examples. In addition to the three classical spacetime theories already discussed, the following examples provide an intuitive starting point for our investigation into what this relationship of 'having more structure than' amounts to.

Example 1. A topological space (X, τ) has more structure than a set X. It has topological structure τ in addition to the basic set structure of X.

Example 2. An inner product space $(V, \langle -, - \rangle)$ has more structure than a vector space V. It has the inner product structure $\langle -, - \rangle$ in addition to all of the basic vector space structure of V.

Example 3. A Riemannian manifold (M, g_{ab}) has more structure than a smooth manifold M. It has the metric structure g_{ab} in addition to all of the manifold structure of M.

These examples are uncontroversial. They are exactly the kinds of examples that are standardly cited as instantiating the 'has more structure than' relation. But not all cases of structural comparison are this straightforward, so we would like to have a general method of determining when one mathematical object has more or less structure than another.

A particularly natural method of comparing amounts of structure has been suggested by both Earman (1989) and North (2009). We will call this method the *automorphism approach*, since it appeals to the automorphisms, or symmetries, of the mathematical objects in question. An automorphism of a mathematical object X is a bijective structure preserving map from X to itself. There is a long tradition in mathematics and physics of looking to the automorphisms of an object for insight into the object's structure. Weyl (1952, p. 144–5) writes, for example, that 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." The automorphism approach is based on precisely this idea. We should look to the automorphisms of objects in order to tell how much structure they have.

Since the automorphisms of an object bear such a close relationship to the structure of the object, one is led to the following kind of criterion for comparing amounts of structure. We use the notation $\mathrm{Aut}(X)$ to denote the group of automorphisms of a mathematical object X.[4]

SYM. A mathematical object X has more structure than a mathematical object Y if and only if the automorphism group $Aut(X)$ is 'smaller than' the automorphism group $Aut(Y)$.

The basic idea behind SYM is clear. If a mathematical object has more automorphisms, then it intuitively has less structure that these automorphisms are required to preserve. Conversely, if a mathematical object has fewer automorphisms, then it must be that the object has more structure that the automorphisms are required to preserve. The amount of structure that a mathematical object has is, in some sense, inversely proportional to the size of the object's automorphism group. Earman (1989, p. 36) puts this basic idea as follows: "As the space-time structure becomes richer, the symmetries become narrower." And North (2009, p. 87) writes that "stronger structure [...] admits a smaller group of symmetries."

But SYM is not useful until we spell out precisely what it means for one automorphism group to be 'smaller than' another. Swanson and Halvorson (2012) and Barrett (2015a, 2015b) have suggested the following way of making this idea precise.

SYM*. *A mathematical object X has more structure than a mathematical object Y if and only if* $\mathrm{Aut}(X) \subsetneq \mathrm{Aut}(Y)$.

The condition $\mathrm{Aut}(X) \subsetneq \mathrm{Aut}(Y)$, i.e. that $\mathrm{Aut}(X)$ is a proper subset of $\mathrm{Aut}(Y)$, is one way to make precise the idea that $\mathrm{Aut}(X)$ is 'smaller than' $\mathrm{Aut}(Y)$. There are two simple arguments that proponents of SYM* give in its favor: the argument from examples and the argument from size. Neither of these arguments, however, are entirely compelling. There is a third argument for SYM*, the argument from definability, that is more involved and given substantially less often. We will discuss these three arguments in turn.

2.1 | The argument from examples

The first argument is an appeal to examples. SYM* makes intuitive verdicts in easy cases of structural comparisons like Examples 1–3, so we should expect it to make the correct verdict in more difficult cases too. In Example 1, for instance, we see that every homeomorphism from the topological space (X, τ) to itself is trivially a bijection from the set X to itself. This means that all automorphisms of the former are automorphisms of the latter. But in general the converse does not hold; there are bijections $X \to X$ that are not homeomorphisms from (X, τ) to itself.

One reasons in a perfectly analogous manner to show that SYM* makes intuitive verdicts in other simple cases as well. For example, one can show that it judges each of the classical spacetime theories discussed above to have less structure than its predecessors (Barrett, 2015b). The fact that SYM* captures some easy examples speaks in favor of the criterion, but it is not entirely convincing. In particular, one would like some kind of conceptual explanation of why SYM* is capturing facts about structural comparison.

2.2 | The argument from size

The second argument in favor of SYM* takes a step in this direction. In fact, we have already seen this argument in the gloss that we gave on SYM earlier. If $\mathrm{Aut}(X) \subsetneq \mathrm{Aut}(Y)$, this means that X has *fewer* automorphisms than Y, which — since automorphisms are structure preserving maps from an object to itself — suggests that X has *more* structure that these automorphisms are required to preserve. This argument is more convincing that the argument from examples. It gives us a kind of explanation for why we should expect SYM* to make reasonable verdicts in cases of structural comparison.

2.3 | The argument from definability

But one can do better. There is a third argument for SYM* that is based on considerations from model theory, and in particular, the theory of definability. These considerations are gestured at by Swanson and Halvorson (2012), but one can make their point more precise. This argument is more convincing, and certainly more illuminating, than either of the previous arguments for SYM*, so it is worth going through it in detail. The remainder of this section will be devoted to doing this.

The simple examples presented above suggest a particularly natural place to start looking for a method for comparing amounts of structure. Consider the following desideratum.

Desideratum. *A mathematical object X has more structure than a mathematical object Y if and only if X has all of the structures that Y has, but X has some structure that Y lacks.*

The idea behind this desideratum is simple, and it coheres well with what is going on in the examples discussed above. Indeed, in each of these cases one can simply read off from the notation that the one object has more structure than the other according to the desideratum. An inner product space $(V, \langle -, - \rangle)$, for example, has all of the structures that a vector space V has, and it has the additional structure provided by the inner product $\langle -, - \rangle$. The notation that we use in these cases allows us to easily point to the additional piece of structure that the one object has. This kind of desideratum is not novel. North (2009, p. 65–66), for example, suggests that one object has more structure than another when the former has additional 'levels' of structure. This idea is essentially the same as our desideratum.

The process of determining whether or not a particular object has a certain piece of structure, however, is not always as simple as just examining notation. Indeed, the notation we use for mathematical objects is not always perfectly transparent with respect to structure. We cannot always read off from the notation we use whether or not X has all of the structures that Y has. Consider the following two examples.

Example 4. A metric space (X, d) has more structure than a topological space (X, τ). At first glance, the notation suggests that a topological space has a piece of structure that a metric space lacks in the form of the topology τ. It is well known, however, that a metric space naturally comes equipped with — indeed, the metric d determines — a canonical topology τ_d. So despite the fact that it cannot be read off simply by looking at the notation, there is a clear sense in which a metric space (X, d) has all of the structures that a topological space has, and in addition, it has metrical structure d that the topological space lacks.

Example 5. A Riemannian manifold (M, g_{ab}) has more structure than a manifold with only a derivative operator (M, ∇). Once again it might appear as if a Riemannian manifold does not have all of the structures that a manifold with derivative operator has; the derivative operator (or 'affine connection') ∇ is not explicitly appealed to in the notation that we use for a Riemannian manifold. It is nonetheless well known that a Riemannian manifold comes equipped with a canonical derivative operator, the Levi-Civita derivative operator. Once again, it cannot be read off of notation, but there is nonetheless a strong sense in which a Riemannian manifold has all of the structures that a manifold with a derivative operator has and more.

The phenomenon illustrated in these two examples is already familiar to mathematicians and logicians. In each case, although the notation we use for the latter object explicitly appeals to some structures that are not among the structures explicitly appealed to in the notation used for the former object, the structures of the former object suffice to *define* all of the structures of the latter object. The topological space (X, τ) has the topological structure τ, which is not among the structures appealed to in the notation we use for the metric space (X, d). But the metric d suffices to define the metric topology τ_d in a perfectly standard manner. Because of this, it is standard to say that the metric space (X, d) 'determines' or 'comes equipped with' or 'naturally gives rise to' topological structure. The same holds in Example 5. Since the metric g_{ab} defines the Levi-Civita derivative operator ∇ on that manifold, it is standard to think of the Riemannian manifold (M, g_{ab}) as determining or coming equipped with a derivative operator.[5]

These two examples show that in general a mathematical object comes equipped with some structures that are not explicitly appealed to in its notation. In fact, it is natural to think of an object as coming equipped not just with its 'basic level' of structure, but also with all of those structures that the basic level defines. With this idea in mind, we can take a step towards sharpening our original desideratum.

Desideratum. *A mathematical object X has more structure than a mathematical object Y if and only if X can define all of the structures that Y has, but X has some piece of structure that Y does not define.*

With this sharpened desideratum in hand, one can ask whether or not SYM* satisfies it. The following question does just this.

Question 1. Is it the case that X has more structure than Y according to SYM* if and only if X defines all of the structure that Y has, but X has some piece of structure that Y does not define?

The argument from definability is an argument for the claim that SYM* does satisfy the desideratum. It answers Question 1 in the affirmative.

At this point, we still need to make precise what it means for X to 'define all of the structures' that Y has. The concept of definability is well-understood in the context of model theory. We therefore need some basic preliminaries on this framework.[6] 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 **Σ-sentence** is a Σ-formula with no free variables. An **automorphism** of a Σ-structure A is a bijection from A to itself that preserves the extensions of all of the predicates, functions, and constants in Σ.

The basic set-up that we will employ in order to discuss definability is the following:

- Let Σ_1 and Σ_2 be signatures. The elements of Σ_1 and Σ_2 represent the 'basic structures' on the two objects that we will consider. These can be thought of as the structures that are explicitly appealed to in the notation we use to describe the objects.
- Let A be a Σ_1-structure and B a Σ_2-structure. We will think of A and B as the two objects (that is, X and Y from our desideratum) whose structures will we be comparing. In order to simplify matters here, we will assume that A and B both have the same underlying set.[7]

We need to make precise what it means for A to define all of the basic structures that B has, or in other words, what it means for A to define all of the elements of Σ_2. So let $p \in \Sigma_2$ be one of the basic structures on B. We assume without loss of generality that p is a predicate symbol. There are now two particularly natural ways to make precise what it means for A to define this additional piece of structure p. We will consider the following two. We say that the Σ_1-structure A **explicitly defines** p^B if there is a Σ_1-formula ϕ such that $\phi^A = p^B$. And we say that the Σ_1-structure A **implicitly defines** p^B if $h[p^B] = p^B$ for every automorphism $h : A \to A$ of A.[8]

The intuition behind these two notions of definability is easy to appreciate. If A explicitly defines the structure p^B, this shows that the structure p^B can be 'constructed from' the basic structures in Σ_1 that A is equipped with. Indeed, it is natural to think of p^B as an 'abbreviation' of the structure ϕ^A that A already has. On the other hand, suppose that A implicitly defines p^B. When this is the case, one often says that the structure p^B is 'invariant under' or 'preserved by' the symmetries of A. And it is common to infer from this that A comes equipped with the structure

p^B.[9] The relation between these two varieties of definability is already well known. If A explicitly defines p^B, then A implicitly defines p^B. But the converse does not hold.

We now have two varieties of definability on the table, so we have a choice about exactly how to make Question 1 precise. We begin by considering the version that asks about implicit definability. We will return to the explicit definability version in the following section.

Question 1 (Implicit definability)**.** Is it the case that X has more structure than Y according to SYM* if and only if X implicitly defines all of the structure that Y has, but X has some piece of structure that Y does not implicitly define?

One can easily prove the following proposition, which shows that that the answer this version of Question 1 is yes.

Proposition 1. *The following are equivalent:*

1. For every symbol $p \in \Sigma_2$, A implicitly defines p^B, but there is a $q \in \Sigma_1$ such that B does not implicitly define q^A.
2. $\text{Aut}(A) \subsetneq \text{Aut}(B)$

Proof. Immediate from definitions. □

The second statement in Proposition 1 says that A has more structure than B according to SYM*; the first statement says that A has more structure than B according to our desideratum, once sharpened using the concept of implicit definability. So Proposition 1 illustrates that SYM* is in fact equivalent to our desideratum in the first-order case, answering Question 1 in the affirmative. This is the argument from definability. It can be summarized as follows. Once one makes it precise using the concept of implicit definability, our desideratum — a particularly natural starting place when looking for a criterion to compare amounts of structure — is the same thing as SYM*.

3 | PROBLEMS WITH THE AUTOMORPHISM APPROACH

The three arguments in the previous section demonstrate that SYM* is a plausible criterion for comparing amounts of structure. And indeed, it has already been fruitfully used to clarify the relationships between different spacetime theories (Barrett, 2015b; Weatherall, 2019a; Bradley, 2020) and between different formulations of classical mechanics (Barrett, 2015a; Swanson & Halvorson, 2012). But the criterion nonetheless suffers from two shortcomings.

3.1 | Sensitivity

The first problem is easy to appreciate.[10] SYM* is too sensitive to the underlying sets of the objects being compared. The following simple example illustrates this problem.

Example 6. Let (X, τ) be a topological space and Y a set that is not equal to X. One wants to say that (X, τ) has more structure than Y. It has topological structure that Y does not have. But it is easy to verify that SYM* does not make this verdict. In fact, no automorphism of (X, τ) is an automorphism of Y and no automorphism of Y is an automorphism of (X, τ). This is because an automorphism of (X, τ) is a bijection from X to itself, and therefore, since X and Y are different sets, not even a function from Y to itself. So according to SYM* these two objects have incomparable amounts of structure. They clearly do not have the same structure, but neither has more nor less structure than the other.

This verdict is troubling, and one can easily come up with many other examples of this same kind. Examples like this show that there is a sense in which SYM* is too *strict* a criterion for comparing amounts of structure. There are pairs of objects X and Y — like the ones from Example 6 — such we want to say that X has more structure than Y, but SYM* does not make this verdict.

At the very least, the sensitivity problem shows us that SYM* is only a useful tool for comparing structure when the objects under consideration have the same underlying set. While there are cases of interest, like many spacetime theories, where the objects under consideration do have the same underlying set, there are also cases of interest where this condition does not hold. The statespace of Lagrangian mechanics is the tangent bundle T_*M, while the statespace of Hamiltonian mechanics is the cotangent bundle T^*M. These objects do not have the same underlying set. Similarly, a relativistic spacetime and an Einstein algebra do not have the same underlying set. But we nonetheless want to be able to compare the structures of these objects. SYM* is not a tool that allows us to do so.

3.2 | Triviality

The second problem is of a slightly different character. SYM* makes implausible verdicts when presented with objects whose only automorphism is the identity map. In such cases we say that the object has a **trivial automorphism group**. The following example makes this problem precise.

Example 7. Let $\Sigma_1 = \{c_1, c_2, c_3, ...\}$ be a signature containing a countable infinity of constant symbols, and let $\Sigma_2 = \Sigma_1 \cup \{p\}$, where p is a unary predicate symbol. We define a Σ_1-structure A and a Σ_2-structure B in the following manner. We let the domains of A and B both be the set $\{0, 1, 2, ...\}$ and we let $c_i^A = c_i^B = i$ for each i. Now note that since there are uncountably many subsets of A, but only countably many Σ_1-formulas, there must be some subset of A that is not equal to ϕ^A for any Σ_1-formula ϕ. We let p^B be one such subset.

There is a sense in which B has more structure than A. Indeed, B is obtained from A by 'adding' a piece of structure in the form of the predicate symbol p. Moreover, because of how we constructed the piece of structure p^B, it is not explicitly definable in terms of the structures on A. But SYM* does not consider B to have more structure than A. Since every automorphism h of A satisfies $h[c_i^A] = c_i^A$, the only automorphism of A is the identity map. So A has a trivial automorphism group. It therefore cannot be the case that the automorphism group of B is a proper subset of the automorphism group of A. This means that it is not the case that B has more structure than A according to SYM*.

In brief, the triviality problem that SYM* faces is the following: When an object X has a trivial automorphism group, there is no object that has more structure than X according to SYM*. Even if an object Y is constructed by adding a level of structure to X — as is the case in the Example 7 — Y will not have more structure than X according to SYM*. That is troubling.

It is worth taking a moment to further unravel Example 7. This example demonstrates a sense in which SYM* actually *does not* perfectly satisfy our desideratum, pointing to a flaw in the argument from definability. Consider again the objects A and B. According to the desideratum, it should be that B has more structure than A. B trivially has all of the structure that A has, *and* it has more, in the form of the predicate symbol p^B that is not definable from the structures on A. But SYM* does not make this verdict. The proper diagnosis of this situation is the following. When the argument from definability — and in particular, Proposition 1 — answered Question 1 in the affirmative, we were employing the notion of implicit definability. Proposition 1 demonstrates that X has more structure that Y according to SYM* if and only if X implicitly defines all of the structures that Y has, but X has some piece of structure that Y does not implicitly define. But this variety of implicit definability is a particularly weak kind of definability. Indeed, if an object X has a trivial automorphism group, every new piece of structure is implicitly definable from the basic structures on the object, since every piece of structure is invariant under the identity map.

Explicit definability is a stronger kind of definability. One might hope that Question 1 can be answered in the affirmative when it is made precise in the following way.

Question 1 (Explicit definability). Is it the case that X has more structure than Y according to SYM* if and only if X explicitly defines all of the structure that Y has, but X has some piece of structure that Y does not explicitly define?

Unfortunately, the answer to this question is no, and Example 7 demonstrates precisely this. The object B explicitly defines all of the structures that A has, A does not explicitly define the structure p^B that B has, but nonetheless it is not the case that B has more structure than A according to SYM*.[11] The triviality problem therefore shows us that while the argument from definability goes through if one is employing the weak notion of implicit definability, it does not go through if one is employing the stronger notion of explicit definability.

We therefore have the following precise diagnosis of what is generating the triviality problem. SYM* is a tool for tracking facts about implicit definability, but it does not perfectly track facts about explicit definability. And this is a definite shortcoming of SYM* as a criterion for comparing amounts of structure. One might hope for a criterion that is more closely connected to explicit definability.[12]

4 | THE CATEGORY APPROACH

Fortunately, there is another approach to comparing amounts of structure that is already on the table: the category approach.[13] This approach involves a shift of emphasis from the automorphism approach. Criteria like SYM* tell us how to compare amounts of structure between individual objects X and Y. There is a sense in which the category approach changes the question. It instead tries to capture when one *type* of mathematical object has more structure than another *type* of mathematical object or when one *theory* posits more structure than another *theory*.

This section will examine one particular way to use categories to compare structures between theories that has recently received attention from logicians and philosophers of physics. In order to explain this method of comparing amounts of structure, we need some basic category theoretic machinery.[14] The class of models of a theory often has the structure of a category. We will call this category the **category of models** of the theory. A **category** C is a collection of objects with arrows between the objects that satisfy two basic properties. First, there is an associative composition operation $\circ$ defined on the arrows of C, and second, every object c in C has an identity arrow $1_c : c \to c$. Let C and D be categories. A **functor** $F : C \to D$ is a map from objects and arrows of C to objects and arrows of D that satisfies

$$F(f : a \to b) = Ff : Fa \to Fb \qquad F(1_c) = 1_{Fc} \qquad F(g \circ h) = Fg \circ Fh$$

for every arrow $f : a \to b$ in C, every object c in C, and every composable pair of arrows g and h in C. Functors are the 'structure preserving maps' between categories; they preserve domains, codomains, identity arrows, and the composition operation. One can think of them as 'translations' between categories; they map objects and arrows of one category to objects and arrows of the other. One property that a functor might have will be crucial in what follows. 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 $Ff = g$.

Baez, Bartels, Dolan, and Corfield (2006) classify functors between categories based on 'what they forget.' Most important for our purposes is the following.

The Baez method. *Let C and D be categories with $F : C \to D$ a functor. We say that F* ***forgets structure*** *if F is not full.*

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. So in order to say whether objects in C have more or less structure than objects in D, the Baez method advises us to look to the kinds of functors that exist between these two categories. Given two theories T_1 and T_2, we look to the functors that exist between their categories of models Mod(T_1) and Mod(T_2) to tell whether one of the theories posits more or less structure than the other. This method of comparing amounts of structure has recently been used widely in philosophy of physics. For example, Feintzeig (2017) uses it to compare structures in quantum theory, Rosenstock et al. (2015) use it to compare formulations of general relativity, Rosenstock and Weatherall (2016), Nguyen et al. (2017), and Weatherall (2016b, 2017a) apply it to different gauge theories, and Barrett (2017) uses it to compare structures in classical mechanics.[15]

The three arguments that we gave in favor of SYM* all translate into arguments in favor of the Baez method. As before, arguments from examples and size are found more often in the literature, but they are less compelling than the argument from definability. Our next aim is to go through these arguments in detail, just like we did for SYM*.

4.1 | The argument from examples

Recall our simple examples from above. One can use the Baez method to capture the structural relationship in each of these cases.

Example 1 (continued)**.** 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 : Top → 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 is not full and therefore forgets structure.

Example 4 (continued)**.** The category Met of metric spaces contains objects (X, d), where d is a metric on the set X. The arrows in Met are isometries, i.e. functions f between metric spaces that preserve the metric d. There is a functor $V : \text{Met} \to \text{Top}$ defined by

$$V : (X, d) \longmapsto (X, \tau_d) \qquad\qquad V : f \longmapsto f$$

where τ_d is the metric topology on X, i.e. the topology that contains all sets that can be realized as unions of open balls according to d. One can easily verify that V is a functor. It is not full, and therefore forgets structure, since there are in general continuous maps between metric spaces that are not isometries.

One can easily verify that the same holds of the other examples discussed earlier. It is also worth seeing a case where the Baez method says that structure has not been forgotten. General relativity can be formulated either using the formal apparatus of a manifold with metric of signature (1,3) or a manifold with metric of signature (3,1). These two formulations only differ with respect to a choice of sign convention, so there is a strong sense in which they ascribe precisely the same amount of structure to spacetime. And indeed, the following example shows that the Baez method makes exactly this verdict.[16]

Example 8. We define a category GR_1 corresponding to the former formulation. An object in GR_1 is a pair (M, g_{ab}) where M a smooth manifold and g_{ab} is a metric on M of signature (1,3). An arrow in GR_1 is a smooth map between manifolds that preserves the metric. Similarly, the category GR_2 has as objects pairs (M, g_{ab}), where M is a smooth manifold and g_{ab} is a metric of signature (3,1). An arrow in GR_2 is again a smooth map between manifolds that preserves the metric.

Consider the functor $F : \text{GR}_1 \to \text{GR}_2$ defined as follows.

$$F : (M, g_{ab}) \longmapsto (M, -g_{ab}) \qquad\qquad F : f \longmapsto f$$

It is easy to check that F is indeed a functor between these two categories. And furthermore, it is easy to verify that F is full.

The Baez method makes intuitive verdicts in many other simple cases too. This shows that the argument from examples translates into an argument in favor of the Baez method too: It makes the intuitive verdicts in many easy cases of structural comparison.

4.2 | The argument from size

Furthermore, the general motivation behind the category approach is essentially the same as that behind automorphism approach. This means that the argument from size carries over to an argument in favor of the Baez method too.

Consider again the case of sets and topological spaces. 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. Roughly, some of the arrows between some sets are not in the 'image' of U. So there are in general 'more' functions between the underlying sets X and Y than there are continuous maps between the topological spaces (X, τ_1) and (Y, τ_2). The arrows in these categories are structure preserving maps between the objects. 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.

The idea behind the argument from size for SYM* was that a larger automorphism group should indicate that the object has less structure. The idea behind the argument from size for the Baez method is perfectly analogous: a larger collection of arrows in a category should indicate that the objects in the category have less structure.

4.3 | The argument from definability

Since both the argument from examples and the argument from size translate into arguments for the Baez method, one naturally wonders whether the argument from definability does too. The aim of this section is to show that it does. This is important because, as with the automorphism approach, neither the argument from examples nor the argument from size is entirely compelling. The basic idea behind the argument from definability is the following. A full functor $F : C \to D$ captures a sense in which the objects of C and D can define one another's structures. And when F is not full, that captures a sense in which the objects of D do not define some of the structure that objects of C have.

In order to demonstrate this, we need some further preliminaries. A Σ-**theory** T is a set of Σ-sentences. A Σ-structure M is a **model** of the Σ-theory T if $M \vDash \phi$ for all $\phi \in T$. A Σ-theory T **entails** a Σ-sentence ϕ, written $T \vDash \phi$, if $M \vDash \phi$ for every model M of T. An **elementary embedding** $f : M \to N$ between Σ-structures M and N is a function from M to N that satisfies

$$M \vDash \phi[a_1, \ldots, a_n] \text{ if and only if } N \vDash \phi[f(a_1), \ldots, f(a_n)]$$

for all Σ-formulas ϕ and elements $a_1, \ldots, a_n \in M$. The collection of models of a Σ-theory T has the structure of a category. We will use the notation $\text{Mod}(T)$ to denote the **category of models** of T. An object in $\text{Mod}(T)$ is a model M of T, and an arrow $f : M \to N$ between objects in $\text{Mod}(T)$ is an elementary embedding $f : M \to N$ between the models M and N. One can easily verify that $\text{Mod}(T)$ is a category.

Recall the desideratum that we appealed to in the automorphism approach. We want to capture a sense in which the Baez method also satisfies that kind of desideratum. In order to be precise, however, we need to replace the talk of individual objects in the earlier desideratum with talk of theories, since it is the structure of entire theories, rather than individual objects, that the Baez method is better equipped to compare. This replacement yields the following desideratum.

Desideratum. *A theory T_2 posits more structure than a theory T_1 if and only if T_2 defines all of the structures of T_1, but T_2 posits some piece of structure that T_1 does not define.*

In order for the Baez method to satisfy this desideratum, it must be that a full functor between categories of models of first-order theories witnesses that a kind of definability relation holds between the two theories and that a non-full functor witnesses that this definability relation does *not* obtain. In particular, we would like the answer to the following question to be yes.

Question 2. Does the existence of a functor $F : Mod(T_2) \to Mod(T_1)$ that forgets structure indicate that T_2 defines all of the structures of T_1, but T_2 posits some piece of structure that T_1 does not define?

This question is asking whether or not the desideratum holds of the Baez method. For our purposes, we will focus on a class of particularly 'well-behaved' functors between categories of models. The general case is beyond the scope of this paper.[17] In order to answer Question 2 we will proceed in two steps. First, we need to describe these well-behaved functors. Intuitively, they are those functors that are induced by syntactic translations between the underlying theories. And second, we present our main result, which isolates what the fullness of a functor is saying about its underlying translation and allows us to answer Question 2 in the affirmative.

We can roughly describe the special kind of functor that we will be considering in the following manner. We most often present a functor $F : C \to D$ by laying down a 'recipe' for how to construct objects of D out of objects of C. For example, the functor from Example 1 provides a recipe for how to construct a set out of a topological space: We simply take the underlying set of the topological space. Similarly, the functor from Example 8 provides a recipe for how to construct a manifold with metric of signature (3,1) out of a manifold with metric of signature (1,3): We take the same underlying manifold and multiply the metric by -1.

This method of defining a functor by providing a recipe is similar to the process of specifying a 'translation' from D to C. Consider Example 4, where we have a functor from the category of metric spaces to the category of topological spaces. In this case, the recipe tells us to take the same underlying set as our metric space, forget about the metric, but leave the metric topology. This recipe goes hand-in-hand with a translation from the 'language of topological spaces' into the 'language of metric spaces'. We know how to talk about topology using only the apparatus of the metric d, and this is what allows us to define the metric topology in the first place. We can, for example, express topological statements like "the function f is continuous" or "the set O is open" using only the metric. In other words, we know how to translate all 'topological talk' into talk of the metric d. The well-behaved functors $F : C \to D$ that we will be considering for the rest of this section have this same feature. They are associated with translations from the 'language of D' to the 'language of C'. It takes a moment to precisely define this special kind of functor.

Let Σ_1 and Σ_2 be signatures, and for simplicity assume that they only contain predicate symbols. A **reconstrual** F of Σ_1 into Σ_2 is a map from the elements of the signature Σ_1 to Σ_2-formulas that takes an n-ary predicate symbol $p \in \Sigma_1$ to a Σ_2-formula $Fp(x_1, \dots, x_n)$ with n free variables.[18] A reconstrual $F : \Sigma_1 \to \Sigma_2$ extends to a map from arbitrary Σ_1-formulas to Σ_2-formulas in the usual recursive manner. In the case where one is only considering signatures with predicate symbols (as we are here), this map is particularly easy to describe. Let $\phi(x_1, \dots, x_n)$ be a Σ_1-formula. We define the Σ_2-formula $F\phi(x_1, \dots, x_n)$ recursively as follows.

1. If $\phi(x_1, \dots, x_n)$ is $x_i = x_j$, then $F\phi(x_1, \dots, x_n)$ is the Σ_2-formula $x_i = x_j$.
2. If $\phi(x_1, \dots, x_n)$ is $p(x_1, \dots, x_n)$, where $p \in \Sigma_1$ is an n-ary predicate symbol, then $F\phi(x_1, \dots, x_n)$ is the Σ_2-formula $Fp(x_1, \dots, x_n)$.
3. If $F\phi$ and $F\psi$ have already been defined for Σ_1-formulas ϕ and ψ, then we define the Σ_2-formula $F(\neg\phi)$ to be $\neg F\phi$, $F(\phi \wedge \psi)$ to be $F\phi \wedge F\psi$, $F(\forall x\phi)$ to be $\forall xF\phi$, etc.

If T_1 and T_2 are theories in the signatures Σ_1 and Σ_2, then we say that a reconstrual $F : \Sigma_1 \to \Sigma_2$ is a **translation** $F : T_1 \to T_2$ if $T_1 \vDash \phi$ implies that $T_2 \vDash F\phi$ for every Σ_1-sentence ϕ. A translation F gives rise to a map $F^* : \text{Mod}(T_2) \to \text{Mod}(T_1)$, which takes models of the theory T_2 to models of the theory T_1. For every model A of T_2 we first define a Σ-structure $F^*(A)$ as follows.

- $\text{dom}(F^*(A)) = \text{dom}(A)$.
- $(a_1, \dots, a_n) \in p^{F^*(A)}$ if and only if $A \vDash Fp[a_1, \dots, a_n]$.

One can show that M and $F^*(M)$ are related to one another in the following way. One uses this lemma to show that $F^*(A)$ is indeed a model of T (Barrett & Halvorson, 2016a, §4).

Lemma. *Let M be a model of T_2 and $\phi(x_1, \dots, x_n)$ a Σ_1-formula. Then $M \vDash F\phi[a_1, \dots, a_n]$ if and only if $F^*(M) \vDash \phi[a_1, \dots, a_n]$.*

The map F^* naturally extends to a mapping on elementary embeddings so that $F^* : \text{Mod}(T_2) \to \text{Mod}(T_1)$ is a functor between the categories of models of T_2 and T_1. If $f : M \to N$ is an arrow between models of T_2, then we define $F^*(f) = f$. One uses the Lemma to verify that $F^*(f)$ is an elementary embedding. Altogether, this means that a translation $F : T_1 \to T_2$ gives rise to a functor $F^* : \text{Mod}(T_2) \to \text{Mod}(T_1)$.

These functors $F^* : \text{Mod}(T_2) \to \text{Mod}(T_1)$ that are induced by translations $F : T_1 \to T_2$ are the special kind of well-behaved functors that we will consider. The translation $F : T_1 \to T_2$ can be thought of as the 'recipe' that gives rise to the functor. Now that we have described these well-behaved functors, we can pose the following 'special case' of Question 2 that we want to consider.

Question 2 (well-behaved functors). Does the existence of a functor $F^* : Mod(T_2) \to Mod(T_1)$ that forgets structure indicate that T_2 defines all of the structures of T_1, but T_2 posits some piece of structure that T_1 does not define?

In order to answer this question in the affirmative, we need to isolate what the fullness of F^* is telling us about the underlying translation F. We say that a translation $F : T_1 \to T_2$ is **essentially surjective** if for every Σ_2-formula ψ there is a Σ_1-formula ϕ such that

$$T_2 \vDash \forall x_1 \dots \forall x_n(\psi(x_1, \dots, x_n) \leftrightarrow F\phi(x_1, \dots, x_n))$$

The existence of an essentially surjective translation $F : T_1 \to T_2$ captures a sense in which T_1 can define all the structures of T_2, since any formula ψ in the language of T_2 is expressible using the language of T_1. The essential surjectivity of F guarantees that there is some formula ϕ in the language of T_1 that translates to (a logical equivalent of) ψ. Intuitively, this means that the theory T_1 can construct all of the concepts that are expressible in the language of T_2. If one thinks of the 'ideology' of a theory as the range of concepts that are expressible in the language in which the theory is formulated (Quine, 1951), then the existence of an essentially surjective translation $F : T_1 \to T_2$ captures a sense in which T_1 and T_2 are as ideologically rich as one another.[19]

For simplicity we assume that the signatures Σ_1 and Σ_2 contain only predicate symbols and are disjoint. We have the following main result linking the fullness of F^* to the essential surjectivity of the translation F.

Proposition 2. *Let T_1 be a Σ_1-theory and T_2 a Σ_2-theory with $F : T_1 \to T_2$ a translation. The following are equivalent:*

1. F is essentially surjective.
2. $F^* : \mathrm{Mod}(T_2) \to \mathrm{Mod}(T_1)$ is full.

It is straightforward to show that 1 implies 2. The opposite direction is more involved and follows from a version of Beth's theorem. Both directions are proven in the appendix.

With Proposition 2 in hand, we can turn back to Question 2 and answer it in the affirmative. The idea is the following. Suppose that F^* is not full and therefore forgets structure according to the Baez method. Since F^* is induced by the translation $F : T_1 \to T_2$ there is a sense in which T_2 can define all of the structures of T_1. For each piece of structure p that T_1 posits, T_2 posits the corresponding piece of structure Fp, and can therefore use this piece of structure to define p.[20] So T_2 defines all of the structures of T_1. And furthermore, since F^* is not full, Proposition 2 guarantees that F is not essentially surjective, and so there is a formula ψ in the language of T_2 — or, in other words, a piece of structure that T_2 posits — for which there is no corresponding piece of structure ϕ posited by T_1 that F translates to ψ. This means that T_1 does not define all of the structures of T_2. On the other hand, when F^* is full and does not forget structure, Proposition 2 guarantees that F is essentially surjective, capturing a sense in which T_1 *does* define the structures of T_2. So we have a strong sense in which the answer to Question 2 is yes. This captures the sense in which the Baez method satisfies our desideratum.

5 | PROBLEMS WITH THE CATEGORY APPROACH

We have seen that all of the arguments in favor of SYM* can be converted into arguments in favor of the Baez method. The Baez method makes intuitive verdicts in many simple cases of structural comparison (the argument from examples), and it is motivated by the simple idea that more symmetries should indicate less structure (the argument from size). And lastly, we have a result showing that, at least when we restrict our attention to well-behaved functors, the Baez method captures a definability relation between the theories under consideration (the argument from definability). It is worth now considering some of the problems that the Baez method faces.

5.1 | Sensitivity and triviality

We first note that the problems of sensitivity and triviality do not pose as much of a threat to the Baez method as they did to SYM*. The problem of sensitivity simply disappears. Unlike SYM*, the Baez method is not concerned with the underlying sets of the objects under consideration.

The problem of triviality is more interesting. Recall how Example 7 generated a problem for SYM*. The objects A and B from that example have the same automorphism group, so according to SYM*, it is not the case that B has more structure than A. But B does come equipped with a piece of structure p^B that A cannot define. There is a sense in which this example generates a problem for the Baez method and a sense in which it does not. The sense in which is does is the following. An automorphism group of an object is trivially a category with a single object; the automorphisms are arrows from that object to itself. And since both of these categories have only one arrow (the identity map on the one object in the category) the functor between $\mathrm{Aut}(B) \to \mathrm{Aut}(A)$ does not forget structure. So according to the Baez method, there is a sense in which B does not have more structure than A. By the letter of the law, therefore, the Baez method makes an undesirable verdict in this case.

The spirit of the law, however, is another matter. There is a sense in which the Baez method is perfectly capable of making the correct verdict. Recall that when we moved from the automorphism approach to the category approach we changed our focus from individual models to theories. It is most natural to think of the Baez method as attempting to tell us when one *theory* posits more structure than another theory. And when we consider the theories that are lurking behind the scenes in Example 7, the Baez method makes a more intuitive verdict.

Example 7 (continued). Let $\mathrm{Th}(B)$ be the $\Sigma_1 \cup \{p\}$-theory that has as axioms every $\Sigma_1 \cup \{p\}$-sentence ϕ such that $B \vDash \phi$, and let $\mathrm{Th}(A)$ be the Σ_1-theory that has as axioms every Σ_1-sentence ψ such that $A \vDash \psi$. Now consider the translation $F : \mathrm{Th}(A) \to \mathrm{Th}(B)$ defined by

$$F : c_i \longmapsto c_i$$

for every constant symbol $c_i \in \Sigma_1$.[21] It is easy to see that F is a translation since for any Σ_1-sentence ϕ, if $A \vDash \phi$, then $B \vDash \phi$. But it is similarly easy to see that F is not essentially surjective. The construction of the predicate p^B guarantees that there is no Σ_1-formula ϕ such that $\mathrm{Th}(B) \vDash \forall x(\phi(x) \leftrightarrow p(x))$, which immediately implies that F is not essentially surjective. So Proposition 2 implies that the functor $F^* : \mathrm{Mod}(\mathrm{Th}(B)) \to \mathrm{Mod}(\mathrm{Th}(A))$ — which takes a model of $\mathrm{Th}(B)$ and 'forgets' about the extension of the predicate p — forgets structure, capturing a sense in which models of $\mathrm{Th}(B)$ like B *do* have more structure than models of $\mathrm{Th}(A)$ like A.

The Baez method is therefore capable of making the intuitive verdict in this case, so long as one takes care to apply it to entire theories like $\mathrm{Th}(A)$ and $\mathrm{Th}(B)$ rather than individual models like A and B. The triviality problem still lingers, but it is less worrying for the Baez method than it was for SYM*.

5.2 | Relativization to the functor

The Baez method avoids the sensitivity and triviality problems by being more 'flexible' than SYM* was. The Baez method is not forced to use the subset relation to judge when one collection of symmetries is larger or smaller than another, and the sensitivity problem is thereby avoided. Similarly, the Baez method is flexible enough to consider the theories corresponding to the objects A and B from Example 7, rather than simply focusing on the objects themselves, and thereby sidesteps the triviality problem.

This flexibility leads, however, to a new issue that is particular to the Baez method. The issue is the following: The choice of functor plays a crucial role in the verdicts that the Baez method makes. Indeed, the Baez method only makes a verdict about which theory posits more structure after one has chosen a functor to use to compare the theories. And in general, there are many functors between two categories. The question is then the following: Which (if any) are the 'right' functors to consider when we compare the structure of objects in category C to objects in category D?

We already stumbled upon a version of this problem when giving the argument from definability. In the case of first-order theories, we can restrict the class of functors that we consider to just the F^* functors — those that arise from syntactic translations between the theories in question — and thereby guarantee that we are only dealing with well-behaved functors. But when two theories are not presented to us in a formal language, like first-order theories are, it is substantially more difficult to say which are the well-behaved functors between their categories of models. Given that these category theoretic tools have recently been used to compare the structures of many physical theories, none of which are formulated using a formal language, one might therefore want to give an account of which functors *in general* are of the well-behaved variety.[22]

The problem actually becomes more pressing. Even if we restrict our attention to well-behaved functors, we can see that different choices of well-behaved functor will lead to different verdicts about which theory posits more structure. The following example illustrates this strange consequence of the Baez method.

Example 9. Consider the two signatures $\Sigma_1 = \{p, q\}$ and $\Sigma_2 = \{r, s\}$, where all of these symbols are unary predicates, and let the Σ_1-theory T_1 and Σ_2-theory T_2 be the empty theories — that is, those with no axioms — in each of these signatures. Now consider the following three translations between these two theories.

$$F : p \longmapsto r \qquad F : q \longmapsto s$$

$$G : p \longmapsto r \qquad G : q \longmapsto r$$

$$H : r \longmapsto p \qquad H : s \longmapsto p$$

One can easily verify that each of $F : T_1 \to T_2$, $G : T_1 \to T_2$, and $H : T_2 \to T_1$ are translations. Since they are translations, we can consider the functors $F^* : \text{Mod}(T_2) \to \text{Mod}(T_1)$, $G^* : \text{Mod}(T_2) \to \text{Mod}(T_1)$, and $H^* : \text{Mod}(T_1) \to \text{Mod}(T_2)$. Now it is easy to verify the following. F is an essentially surjective translation, so Proposition 2 implies that F^* is full and does not forget structure. But neither G nor H is an essentially surjective translation, so Proposition 4 implies that G^* and H^* are not full, and therefore they both forget structure.

Suppose that we ask which of the theories T_1 and T_2 from this example posits more structure. If we use the Baez method to answer this question, the answer we get depends on the functor being considered. The functor F^* does not forget structure; it captures a sense in which the two theories posit the same amount of structure. But the functor G^* does forget structure, capturing a sense in which T_1 posits more structure than T_2; and H^* also forgets structure, capturing a sense in which T_2 posits more structure than T_1. The best way to put this feature of the Baez method is as follows: It only makes a verdict relative to a choice of functor between the two categories. There are many other examples that illustrate this same point.

Addressing this problem with the care it deserves is unfortunately beyond the scope of this paper. It will suffice to gesture at the most natural kind of solution. When presented with two theories we would like a way to identify one particular functor from the one category of models to the other that captures the 'correct' standard of comparison — or translation — between the two theories. This is tantamount to supplementing the Baez method with a procedure for choosing the 'correct' functors to use to compare structure between two theories. One can think of our restriction above to the case of well-behaved functors F^* as a move in this direction. In the case of physical theories, a similar restriction is adopted. It is standard to require that the functors between categories of models for physical theories 'preserve empirical content' — otherwise they would clearly not be suitable translations between the theories (Weatherall, 2016a; Barrett, 2017). This requirement also pares down the options of functors that we might consider.

Although this problem is certainly pressing, it is somewhat assuaged by noticing that the choice of which functors to use to compare two theories is often quite easy. In the case of physical theories there are often particularly 'natural' candidates of functors to choose from (Weatherall, 2019a, 2019b). For example, the Legendre transformation gives rise to the most natural functor between Hamiltonian and Lagrangian mechanics, the famous 'function-space' duality gives rise to the most natural functor between general relativity and the theory of Einstein algebras, and as we saw in Example 8 'flipping the sign' of the metric gives rise to the most natural functor between the (1,3) and (3,1) formulations of general relativity. But of course, merely mentioning that it is often easy to choose a functor to use to compare the structure of two theories is not the same as providing a general method that we can use to make that choice. Further work on the Baez method is therefore necessary.

6 | STRUCTURAL PARSIMONY AND EQUIVALENCE

In the meantime, however, we can draw some conclusions from what we have seen so far. In particular, our discussion yields two payoffs concerning what it is for a theory to 'posit' or 'employ' a particular structure. One has to do with the structural parsimony principle, and the other has to do with equivalence of theories. Both are centered on the idea that a theory might posit or employ some structure that is not *explicitly* appealed to in its formulation. We will take the two payoffs in turn.

6.1 | Structural parsimony

We begin with the structural parsimony principle. One of the reasons we were initially motivated to find a method of comparing amounts of structure was that we wanted to clearly understand the conditions under which the following principle is applicable.

Structural parsimony. All other things equal, we should prefer theories that posit less structure.

In order to justify the principle, one needs to say exactly why theories that posit less structure should be preferred. One reason that is often given is that excising surplus structure from a theory provides us with a more accurate description of reality. The basic idea is that once we excise all of the surplus structures from our theory, we end up with a kind of 'ideal theory' — one that has no arbitrary conventions or surplus structure, and therefore provides us with a window into the 'real structure' of the world. North (2009, p. 78) provides a justification along these lines when she suggests the following method of interpreting our physical theories:

> Take the mathematical formulation of a given theory. Figure out what structure is required by that formulation. [...] Infer that this is the fundamental structure of the theory. Go on to infer that this is the fundamental structure of the world, according to the theory.

Even if we do not think that less structured theories provide us with a more accurate description of the world, there are practical reasons to prefer them. If we use theories whose formulations do not appeal to arbitrary choices of scale, coordinate system, rest frame, etc., then we can avoid mistakenly attributing some representational significance to these conventional aspects of the formulation. More 'intrinsic' theories like this may also offer more illuminating explanations of the phenomena (Field, 2016).

Insofar as one endorses the structural parsimony principle, it is important to know the conditions under which one theory 'does away with' or 'excises' some piece of surplus structure from another theory.[23] Otherwise we would not even know which theories the principle is directing us to prefer. This is where the tools that we have discussed for comparing amounts of structure come into play. The example that is often cited as a 'paradigm case' of successful structure excision is the move from Newtonian to Galilean spacetime.

Example 10. Newtonian spacetime is the tuple $(\mathbb{R}^4, t_{ab}, h^{ab}, \nabla, \lambda^a)$, where $\mathbb{R}^4$ is a smooth manifold, t_{ab} and h_{ab} are the temporal and spatial metrics, ∇ is a derivative operator, and λ^a is a vector field representing the standard of absolute rest. We decided that absolute rest was surplus to the theory, and we were then able to excise it by moving to Galilean spacetime. Galilean spacetime is represented by the tuple $(\mathbb{R}^4, t_{ab}, h^{ab}, \nabla)$. As one can see by examining the notation, the piece of structure λ^a is no longer explicitly referred to or appealed to in our new theory of spacetime. We have successfully excised it from our theory.[24]

Along with many other simple cases of excising structure, this example of classical spacetime theories suggests that all it takes to excise a piece of structure from a theory is to move to a new theory that no longer explicitly appeals to that piece of structure. Galilean spacetime $(\mathbb{R}^4, t_{ab}, h^{ab}, \nabla)$ does not explicitly appeal to the Newtonian standard of rest λ^a in its notation. And one might conclude that it is in virtue of this that Galilean spacetime has excised that structure.

Unfortunately, this is incorrect. It takes more than avoiding explicit appeal to a piece of structure in order to excise it. The following two examples show this. The first is a simple 'toy' case, the second an actual case from physics.

Example 11. Let $\Sigma = \{p, r\}$ be a signature containing the binary predicate p and the unary predicate r. Consider the Σ-theory T with the one axiom $\forall x(r(x) \leftrightarrow p(x,x))$. This theory is simply saying that r is the diagonal of p. Suppose that we want to excise the piece of structure r from the theory T.

Consider the $\{p\}$-theory T^- with no axioms. This theory clearly avoids explicit appeal to the structure r; that symbol appears nowhere in its formulation, nor in any of the sentences that it entails. So if we think that is sufficient to excise a piece of structure, then T^- successfully excises the structure r.

Example 12. General relativity is standardly formulated in terms of a manifold with various geometric structures on it. Geroch (1972) proposed an alternative algebraic formulation of general relativity that has come to be called the *theory of Einstein algebras*. This theory is formulated using an algebra of smooth scalar fields, instead of the standard geometric apparatus of general relativity.

Since the theory of Einstein algebras no longer explicitly appeals to the structure of 'spacetime points' — indeed, its formulation is purely algebraic and does not mention manifolds — Earman (1986, 1989) put forward this theory as a 'relationalist' formulation of general relativity. He suggested that the theory of Einstein algebras excised the structure of spacetime points from general relativity. The basic idea was that the theory of Einstein algebras does to spacetime points what Galilean spacetime does to the Newtonian standard of absolute rest. The structure of spacetime points — and smooth manifold structure in general — is longer explicitly appealed to in the algebraic reformulation of the theory, which only appeals to algebras of smooth scalar fields.[25]

Examples 11 and 12 are analogous to one another. In each case, we attempt to excise a piece of structure by reformulating the theory in such a way that the structure we want to excise is no longer explicitly appealed to. But in neither case is the excision successful. This is because the remaining structures suffice to construct or define the piece of structure that we were trying to excise. We return to both of the examples in turn.

Example 11 (continued)**.** Despite the fact that T^- does not explicitly appeal to the structure r, there is nonetheless a strong sense in which nothing has been excised from T. Indeed, according to all of our methods of comparing amounts of structure, T^- does not have less structure than T. It is easy to verify that a model M of T has precisely the same automorphism group as the corresponding model $M|_{\{p\}}$ of T^-, so according to SYM* no structure has been excised from models of T when we move to T^-. The Baez method makes this same verdict. Consider the functor $F^* : \text{Mod}(T) \to \text{Mod}(T^-)$ associated with the translation $F : T^- \to T$ that maps the predicate symbol p to itself. It is easy to check that F is essentially surjective. Proposition 2 therefore implies that F^* is full and does not forget structure.

The idea behind these results is clear: Even though T^- does not explicitly appeal to the structure r, it is nonetheless definable out of the structures that T^- does explicitly appeal to. So we are justified in saying that T^- does not posit less structure than T, and therefore nothing is excised when we move from T to T^-.

The same holds in the case of general relativity and the theory of Einstein algebras.

Example 12 (continued)**.** Indeed, it has recently been shown by Rosenstock et al. (2015) that the most natural functors between the categories of models for general relativity and the theory of Einstein algebras do not forget structure. One can 'translate' back and forth between the two theories without 'forgetting' anything. So they posit precisely the same amount of structure according to the Baez method.[26] The idea behind this result is once again clear: Even though the theory of Einstein algebras does not appeal to spacetime points explicitly, they are definable out of the structures — namely, the scalar fields — that the theory of Einstein algebras does explicitly appeal to. So in moving from general relativity to the theory of Einstein algebras, there is a sense in which we have excised nothing from our theory.

This example is particularly interesting, since Earman (1986, 1989) famously argued that the theory of Einstein algebras did away with some structure from the standard geometric formulation of general relativity. His idea was the following. It is standard to think that the hole argument indicates that general relativity has some excess structure — though, as Weatherall (2017b) has argued, some of the tools discussed here suggest the contrary. Earman proposed the theory of Einstein algebras as an attempt to excise this structure; he hoped that this new theory would then provide a mathematical setting for a suitably "relationist" formulation of general relativity. The results here — in combination with those of Rosenstock et al. (2015) — illustrate why this does not work. There is a strong sense in which nothing has been excised in the move from general relativity to the theory of Einstein algebras.[27]

These two examples yield a simple payoff that is important to keep in mind, especially insofar as one endorses the structural parsimony principle and aims to excise surplus structure from our theories.

Payoff 1. *Excising a piece of structure from a theory is not as simple as just reformulating the theory in such a way that the piece of structure is no longer explicitly appealed to.*

It can be the case that the structures that are explicitly appealed suffice to define the structure that we were trying to excise. And when that happens, there is a strong sense in which we have actually excised nothing.

The tools that we have discussed in this paper provide us with a guard against making this kind of mistake. We simply have to make sure that when we move to a new theory, we actually have excised *something*. For example, in the case of Newtonian and Galilean spacetime both SYM* and the Baez method agree that the latter posits has less structure than the former (Barrett, 2015b). So something has been excised when we move from Newtonian to Galilean spacetime. And in general, we have shown that SYM* and the Baez method bear a close relationship to definability. So if these methods judge that one theory posits less structure than another, then we have a guarantee that in moving from the former theory to the latter we will not be making the kind of mistake that we made in Examples 11 and 12. In order to have excised something it must be the case that the resulting theory actually posits less structure than the theory that we began with. If the resulting theory does not posit less structure, then we simply have not excised anything *despite* the fact that the resulting theory may not explicitly appeal to the structure that we were trying to excise.[28]

6.2 | Equivalence

Our second payoff has to do with equivalence of theories. It has recently been suggested — by North (2009) and Barrett (2019), among others — that we can use tools for comparing amounts of structure to help us judge whether two theories are equivalent. In particular, if two theories posit different amounts of structure, it must be that they are inequivalent. The Newtonian and Galilean theories of spacetime, for example, are manifestly inequivalent theories. Since they disagree about the amount of structure the spacetime has — the one ascribes the structure of a privileged rest frame to spacetime, the other does not — they are clearly not saying the same thing about the world.

Although this kind of reasoning should work in general — see Barrett (2019) for more detailed discussion — the tools we have discussed here suggest that we must take significant care when applying this reasoning to particular cases. Consider once again the case of general relativity and the theory of Einstein algebras. If we take these two theories at face value, they posit *completely* different structures. General relativity explicitly appeals only to geometric structures — a smooth manifold with metric. The theory of Einstein algebras, on the other hand, explicitly appeals only to algebraic structures. In virtue of the fact that the two theories appeal to completely different kinds of structure, one might be tempted to conclude that they ascribe incomparable amounts of structure to the world. It is not that one of them posits more structure than the other. But rather, each posits some structures that the other does not. If this is so, then we would have good reason to conclude that the two theories are inequivalent.

We do not, however, have good reason to draw this conclusion. The tools that we have discussed here show why. According to the Baez method, these two theories do not posit incomparable amounts of structure. Despite the fact that they explicitly appeal to different structures in their formulations, each theory can *define* the structures of the other. There is a strong sense, therefore, in which they posit exactly the same amount of structure.

We can put this payoff more generally as follows.

Payoff 2. *The fact that two theories explicitly appeal to different collections of structures in their formulations does not imply that they are inequivalent.*

In the case of general relativity and the theory of Einstein algebras, we might expect the two theories to be inequivalent in virtue of positing different amounts of structure. But once we have precise tools for comparing amounts of structure, we can guard against this kind of mistake. The two theories may nonetheless define one another's structures, even though they do not explicitly appeal to the same structures in their formulation.

One might worry about the extent to which these two payoffs depend on SYM* and the Baez method. We have seen that both of these methods of comparing amounts of structure have their faults, so it will be useful to gauge whether these payoffs are sensitive to those methods. Fortunately, we have good reason to think that these two payoffs will still hold even if we move in another direction and do away with both SYM* and the Baez method. If we think that fewer symmetries indicates more structure, then we must be taking definable structure seriously. Adding definable structure does not reduce the symmetry group of an object, so if we like the idea that fewer symmetries is the indication of less structure, then we are also committing to the idea that adding a layer of definable structure is not really adding any structure at all. Rather, it was already there to begin with. Taking definable structure seriously — and not merely taking seriously the structures that are *explicitly* appealed to in a theory's formulation — is what leads to the two payoffs. Insofar as we adopt a method of comparing amounts of structure that, like SYM* and the Baez method, appeals to symmetries, our two payoffs will still hold.

The general thrust of these two payoffs might be summed up as follows: *Reading off the structure that a theory posits is more difficult than just looking to the structures that the theory explicitly appeals to in its formulation*. It is in general not so simple to read off the structure of a theory. The methods of comparing amounts of structure that we have discussed here are tools that can, and indeed should, be used to help.

ACKNOWLEDGMENTS

I'm especially grateful to Hans Halvorson, and Jim Weatherall for many helpful conversations about this material over the years. Thanks to audiences in Munich and Irvine for discussion.

ENDNOTES

[1] See for example Geroch (1978), Friedman (1983), Earman (1989), and Maudlin (2012).

[2] See, for example, Andréka, Madarász, and Németi (2005), Barrett (2017), Barrett and Halvorson (2016a, 2016b, 2017a, 2017b), Coffey (2014), Curiel (2014), Halvorson (2013), Glymour (2013), Hudetz (2015, 2017), Knox (2011, 2014), North (2009), Rosenstock et al. (2015), Rosenstock and Weatherall (2016), Teh and Tsementzis (2017), Tsementzis (2015), Van Fraassen (2014), and Weatherall (2017a). See also the classic work of Glymour (1971, 1977, 1980), Quine (1975), and Sklar (1982). See Weatherall (2019a) for a review of recent work.

[3] See the classic discussions of Earman (1989) and Friedman (1983). Recently these issues have been discussed by North (2009), Swanson and Halvorson (2012), Barrett (2015a, 2015b), Weatherall (2016b), Nguyen, Teh, and Wells (2017), and Feintzeig (2017). Note that depending on exactly what tools we end up with to compare amounts of structure, it may be that structure is not the kind of thing that can be genuinely *counted*; it may be that we cannot represent 'amounts of structure' with the ordering of the real numbers. For example, as we will see, many standards of structural comparison deem it possible for objects to have incomparable amounts of structure, in the sense that neither has more nor less nor the same structure as the other.

[4] The names "SYM" and "SYM*" come from Swanson and Halvorson (2012) and Barrett (2015a, 2015b).

[5] Any topology textbook will give the familiar explicit definition of τ_d in terms of d. Geroch (1972) shows how one can explicitly define ∇ using g_{ab} and the Lie derivative. See Barrett (2018) and the references therein, for a sampling of the rich literature on definability in logic.

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

[7] This assumption seems innocuous at this point — after all, the pairs of objects under consideration in Examples 1–5 all had the same underlying set — but we will return to it later when discussing shortcomings of SYM*.

[8] It is important to note here that there are a number of different varieties of implicit definability that are often considered. The kind of invariance under symmetry condition that we provide here is one of the weaker varieties. The reader is encouraged to consult Barrett (2018), Hodges (2008), and Winnie (1986) and the references therein for further details on implicit and explicit definability.

[9] For discussions about the sense in which this inference is justified, see Barrett (2018) and Dasgupta (2016) and the references therein.

[10] This problem is gestured at in the discussion of a criterion called SYM** by Barrett (2015a), and it is mentioned explicitly by Barrett (2015b, p. 3).

[11] One can easily verify that A does implicitly define p^B, showing, as we remarked earlier without proof, that this variety of implicit definability does not entail explicit definability.

[12] One might just put forward our desideratum itself — made precise using the notion of explicit definability — as a criterion for comparing amounts of structure. The main issue with this is that in order to apply this desideratum to a pair of objects, one would first have to know what "languages" are used to describe the objects. One cannot speak precisely of explicit definability without having a clear picture of what languages the objects are formulated in. Most mathematical objects are not presented to us using a formal language, and this criterion would not be easily applicable in such cases. Our best hope, therefore, is to find a criterion like SYM* (or the categorical criterion that we will discuss shortly) that is easily applicable to mathematical objects "in the wild", but that also bears a close relationship to explicit definability in the cases where the objects under consideration are presented to us in formal languages.

[13] A criterion that replaces the subset relation of SYM* with the subgroup relation, called SYM**, has also been discussed (Barrett, 2015a). The problem with this criterion is that it compares too many different mathematical objects and in doing so actually moves farther away from our desideratum than SYM*.

[14] The reader is encouraged to consult Mac Lane (1971) or Borceux (1994) for further details.

[15] It is also well known in the category theory community. See Baez and Shulman (2010).

[16] See Barrett (2017) for further discussion of this case.

[17] Arbitrary functors between categories of models for first-order theories notoriously do not have the kinds of nice properties that the functors we consider here have. See Barrett and Halvorson (2016b) and especially Hudetz (2017) for examples.

[18] See Hodges (2008), Button and Walsh (2018), Barrett and Halvorson (2016a) for additional details. See Hudetz (2017) for a description of a similar kind of well-behaved functor.

[19] Quine (1951, p. 15) remarks that one can investigate the ideology of a theory by examining what kinds of translations exist between theories: "Much that belongs to ideology can be handled in terms merely of the translatability of notations from one language into another; witness the mathematical work on definability by Tarski and others." Halvorson (2019, p. 120) expresses the same thought.

[20] More precisely, what this means is that there is a reconstrual G from a 'sub-signature' of Σ_2 to Σ_1. Moreover, G is an essentially surjective translation from T_2 (thought of as a theory in this sub-signature) to the extension of T_1 obtained by adding to the axioms of T_1 all of those Σ_1-sentences ϕ such that $T_2 \vDash F\phi$. This essentially surjective translation G is therefore capturing a sense in which T_2, using only some of its structures, can define a special case of T_1's structures.

[21] We did not define above how translations work when the signatures contain constant symbols. The reader is invited to consult Barrett and Halvorson (2016a) for details. Alternatively, one could simply reformulate the two theories here using predicate symbols instead of constant symbols.

[22] See Weatherall (2019a) and Hudetz (2017) for further discussion.

[23] We will not discuss here the difficult question of what makes a piece of structure surplus or superfluous, and thus a *candidate* for excision. That question naturally leads one into the literature on symmetries. See Dasgupta (2015) and the reference therein.

[24] See Barrett (2015b) for further details on this case.

[25] See Rosenstock et al. (2015) and the references therein for further discussion of this example.

[26] SYM* is not equipped to make a verdict in this case since an Einstein algebra and a relativistic spacetime have different underlying sets. Recall the problem of sensitivity.

[27] Roughly this same point was made earlier by Rynasiewicz (1992).

[28] Another way in which one might fail to excise structure is by attempting to excise a piece of structure that actually was not there to excise in the first place. As mentioned earlier, it has recently been suggested by Weatherall (2017b), for example, that something like this happens often in the literature on the famous hole argument in general relativity.

REFERENCES

Andréka, H., Madarász, J. X., & 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., & Corfield, D. (2006). Property, structure and stuff. Available at http://math.ucr.edu/home/baez/qg-spring2004/discussion.html.

Baez, J., & Shulman, M. (2010). Lectures on n-categories and cohomology. In J., Baez & P., May (Eds.) *Towards higher categories*. New York: Springer-Verlag.

Barrett, T. (2019). Structure and equivalence. *Forthcoming in Philosophy of Science*.

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. (2018). What do symmetries tell us about structure? *Philosophy of Science*, *85*, 617–639.

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

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

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

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

Borceux, F. (1994). *Handbook of categorical algebra* (vol. 1). Cambridge University Press.

Bradley, C. (2020). The non-equivalence of Einstein and Lorentz. *Forthcoming in the British Journal for the Philosophy of Science*.

Button, T., & Walsh, S. (2018). *Philosophy and model theory*. Oxford University Press.

Coffey, K. (2014). Theoretical equivalence as interpretative equivalence. *The British Journal for the Philosophy of Science*, *65*(4), 821–844.

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

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.

Earman, J. (1986). Why space is not a substance (at least not to first degree). *Pacific Philosophical Quarterly*, *67*(4), 225–244.

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*.

Field, H. (2016). *Science without numbers*. Oxford University Press, second edition edition.

Friedman, M. (1983). *Foundations of space-time theories: Relativistic physics and philosophy of science*. Princeton University Press.

Geroch, R. (1972). Einstein algebras. *Comm. Math. Phys.*

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

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

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

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

Glymour, C. (2013). Theoretical equivalence and the semantic view of theories. *Philosophy of Science*, *80*(2), 286–297.

Halvorson, H. (2013). The semantic view, if plausible, is syntactic. *Philosophy of Science*, *80*(3), 475–478.

Halvorson, H. (2019). *The logic in philosophy of science*. Cambridge 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*, (pp. 294–308).

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

Knox, E. (2011). Newton-Cartan theory and teleparallel gravity: The force of a formulation. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *42*(4), 264–275.

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

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

Maudlin, T. (2012). *Philosophy of physics: space and time*. Princeton University Press.

Nguyen, J., Teh, N. J., & 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.

Quine, W. V. O. (1951). Ontology and ideology. *Philosophical Studies*, *2*(1), 11–15.

Quine, W. V. O. (1975). On empirically equivalent systems of the world. *Erkenntnis*, *9*(3), 313–328.

Rosenstock, S., Barrett, T. W., & 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., & 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).

Rynasiewicz, R. (1992). Rings, holes and substantivalism: On the program of Leibniz algebras. *Philosophy of Science*, *59*(4), 572–589.

Sider, T. (2013). Against parthood. In K., Bennett, & D. W., Zimmerman (Eds.), *Oxford Studies in Metaphysics*, (vol. 8). Oxford University Press.

Sklar, L. (1982). Saving the noumena. *Philosophical Topics*, pages 89–110.

Swanson, N., & Halvorson, H. (2012). On North’s ‘The structure of physics’. *Manuscript*.

Teh, N., & Tsementzis, D. (2017). Theoretical equivalence in classical mechanics and its relationship to duality. *Forthcoming in Studies in History and Philosophy of Modern Physics*.

Tsementzis, D. (2015). A syntactic characterization of Morita equivalence. *Manuscript*.

Van Fraassen, B. C. (2014). One or two gentle remarks about Hans Halvorson’s critique of the semantic view. *Philosophy of Science*, *81*(2), 276–283.

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*.

Weatherall, J. O. (2019a). Theoretical equivalence in physics. *Forthcoming in Philosophy Compass*.

Weatherall, J. O. (2019b). Why not categorical equivalence? In J., Madarász, & G., Székely (Eds.), *Forthcoming in Hajnal Andreka and Istvan Nemeti on Unity of Science*.

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

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

How to cite this article: Barrett TW. How to count structure. *Noûs*. 2020;1–28. https://doi.org/10.1111/nous.12358

APPENDIX

The purpose of this appendix is to prove Proposition 2, which we restate here for convenience.

Proposition 2. *Let T_1 be a Σ_1-theory and T_2 a Σ_2-theory with $F : T_1 \to T_2$ a translation. The following are equivalent:*

1. F is essentially surjective.
2. $F^* : \text{Mod}(T_2) \to \text{Mod}(T_1)$ is full.

We need a few more definitions before proving this proposition. 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. An **explicit definition** of p in terms of Σ is a Σ^+-sentence of the form

$$\forall x_1 \dots \forall x_n (p(x_1, \dots x_n) \leftrightarrow \phi(x_1, \dots, x_n))$$

where $\phi(x_1, \dots, x_n)$ is a Σ-formula. A **definitional extension** of a Σ-theory T to the signature Σ^+ is a Σ^+-theory

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

such that for each predicate symbol $s \in \Sigma^+ - \Sigma$, the sentence δ_s is an explicit definition of s in terms of Σ. One can easily verify that a definitional extension is indeed an extension. One can also define new function and constant symbols, but for our purposes this will not be important.

When T^+ is an extension of a Σ-theory T, we can define the **projection functor** $\Pi : \text{Mod}(T^+) \to \text{Mod}(T)$ by

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

for every model M of T^+ and elementary embedding h between models of T^+. Here $M|_\Sigma$ is the Σ-structure obtained from M by forgetting the extensions of all the predicates not in Σ. In the case where T^+ is a definitional extension of T, the functor Π is full. Indeed, it is an equivalence of categories (Barrett & Halvorson, 2016b, Propositions 5.1–5.3).

We now have the resources to prove Proposition 2.

Proof that 1 implies 2. Suppose that $F : T_1 \to T_2$ is essentially surjective. Let M and N be models of T_2 with $h : F^*(M) \to F^*(N)$ an elementary embedding. We need to show that $h : M \to N$ is an elementary embedding. So let ψ be a Σ_2-formula. Since F is essentially surjective, we know that there is a Σ_1-formula ϕ such that $T_2 \vDash \forall x_1 \dots \forall x_n(\psi(x_1, \dots, x_n) \leftrightarrow F\phi(x_1, \dots, x_n))$. We then immediately see that the following string of equivalences hold for any elements $a_1, \dots, a_n \in M$:

$$M \vDash \psi[a_1, \dots, a_n] \Leftrightarrow M \vDash F\phi[a_1, \dots, a_n]$$

$$\begin{aligned}
&\Leftrightarrow F^*(M) \vDash \phi[a_1, \dots, a_n] \\
&\Leftrightarrow F^*(N) \vDash \phi[h(a_1), \dots, h(a_n)] \\
&\Leftrightarrow N \vDash F\phi[h(a_1), \dots, h(a_n)] \\
&\Leftrightarrow N \vDash \psi[h(a_1), \dots, h(a_n)]
\end{aligned}$$

The first and fifth equivalences follow from our choice of ϕ, the second and fourth from the Lemma, and the third from the fact that $h : F^*(M) \to F^*(N)$ is an elementary embedding. This implies that $h : M \to N$ is an elementary embedding and so $F^* : \mathrm{Mod}(T_2) \to \mathrm{Mod}(T_1)$ is full. □

Proof that 2 implies 1. Suppose that $F^* : \mathrm{Mod}(T_2) \to \mathrm{Mod}(T_1)$ is full. In order to demonstrate that F is essentially surjective it will suffice to show that for every predicate symbol $q \in \Sigma_2$ there is a Σ_1-formula ϕ such that $T_2 \vDash \forall x(q(x) \leftrightarrow F\phi(x))$. Consider the $\Sigma_1 \cup \Sigma_2$-theory T_2^+ that is defined as follows:

$$T_2 \cup \{\forall x(p(x) \leftrightarrow Fp(x)), p \in \Sigma_1\}$$

T_2^+ is a definitional extension of T_2. Using the fact that F is a translation, one can show that T_2^+ is an extension of T_1. One can then verify using the Lemma that the following diagram commutes, where $\Pi_1 : \mathrm{Mod}(T_2^+) \to \mathrm{Mod}(T_2)$ and $\Pi_2 : \mathrm{Mod}(T_2^+) \to \mathrm{Mod}(T_1)$ are the projection functors.

$$\begin{tikzcd}
\mathrm{Mod}(T_2) \arrow[d, "F^*"] & \mathrm{Mod}(T_2^+) \arrow[l, "\Pi_1"'] \arrow[dl, "\Pi_2"] \\
\mathrm{Mod}(T_1) &
\end{tikzcd}$$

Since T_2^+ is a definitional extension of T_2, Π_1 is full. By assumption F^* is full, so since $\Pi_2 = F^* \circ \Pi_2$ this means that Π_2 must be full too.

Now using the fact that Π_2 is full, Beth's theorem — in particular a simple corollary to it (Barrett, 2018, Corollary 1) — implies that for every predicate symbol $q \in \Sigma_2$ there is a Σ_1-formula ϕ such that $T_2^+ \vDash \forall x(q(x) \leftrightarrow \phi(x))$. We now claim that $T_2 \vDash \forall x(q(x) \leftrightarrow F\phi(x))$. Let M be a model of T_2. We then have the following string of equivalences.

$$\begin{aligned}
a \in q^M \iff a \in q^{M^+} \iff a \in \phi^{M^+} \iff a \in \phi^{\Pi_2(M^+)} &\iff a \in \phi^{F^*(M)} \\
&\iff a \in F\phi^M
\end{aligned}$$

(Here M^+ is the unique model of T_2^+ that satisfies $\Pi_1(M^+) = M$. It is unique since T_2^+ is a definitional extension of T_2.) The first equivalence follows from the definition of Π_1, the second from our choice of ϕ, the third the definition of Π_2, the fourth from the fact that the above diagram commutes, and the fifth from the Lemma. This means that $T_2 \vDash \forall x(q(x) \leftrightarrow F\phi(x))$, so F is essentially surjective. □