Ergo AN OPEN ACCESS JOURNAL OF PHILOSOPHY

What Are Symmetries?

DAVID JOHN BAKER
Department of Philosophy, University of Michigan

I advance a stipulational account of symmetry-to-reality inference, according to which symmetries are part of the content of theories. For a theory to have a certain symmetry is for the theory to stipulate that models related by the symmetry represent the same possibility. I show that the stipulational account compares positively with alternatives, including Dasgupta's epistemic account of symmetry, Møller-Nielsen's motivational account, and so-called formal and ontic accounts. In particular, the stipulational account avoids the problems Belot and Dasgupta have raised against formal and ontic accounts of symmetry while retaining many of the advantages of these otherwise-attractive frameworks.

1. Introduction

A particular sort of inference is common in the philosophy of physics (and in the practice of physics itself, when the question arises). I'm referring to what Shamik Dasgupta (2016) has called a symmetry-to-reality inference: the move from the premise that some quantity or property varies between solutions of a theory related by a symmetry to the conclusion that that quantity or property is unreal according to that theory.[1] Alternatively, we may think of symmetry-to-reality inferences in terms of a theory's models in the following way: if two models of a theory are related by a symmetry, those models represent the same possibility.

I'll begin from the assumption that symmetry-to-reality inferences are reasonable. The challenge I will undertake is to provide an account of what symmetries are that explains why these inferences are reasonable (and that gets sufficiently close to working physicists' use of the term). My proposal is the *stipulational*

1. Although this is how I understand symmetry-to-reality inferences, it is not precisely how Dasgupta understands them. For Dasgupta, only the following sort of symmetry-to-reality inference is valid: "if a putative feature is variant in laws that we have reason to think are true and complete, then this is some reason to think that the feature is not real" (Dasgupta 2016: 840).

Contact: David John Baker <djbaker@umich.edu>

https://doi.org/10.3998/ergo.3592

account of symmetries. On this view, a symmetry is a stipulated relation of representational equivalence between models or solutions of a theory. This means that symmetries are not something to be read off or inferred from the content of a theory; rather, they are part of that content. To state the view precisely: many theories posit (among other things) that certain groups of transformations among their models are *symmetries*, which is to say that two models related by one of those transformations represent the same possibility.

I will begin in Sec. 2 with a brief look at the puzzles surrounding symmetry-to-reality inference. In Sec. 3 I situate my account of symmetry within an overall picture of the interpretation of theories (especially, although not exclusively, physical theories). I emphasize the importance of the fact that we normally interpret non-fundamental theories, beginning from what Ruetsche (2011) calls a partial interpretation. This makes it natural to introduce stipulations about the equivalence of states into our interpretations of theories, and for scientists to include such stipulations in constructing theories. In Sec. 4, I show that the problems of interpretation are fruitfully analogous to certain problems in metaphysics, and that similar stipulations of equivalence can help resolve some of these problems. This also provides a clear illustration of what it means to introduce these stipulations into a theory, and of why they can never appear in a completely fundamental theory of everything. The sense in which these stipulations are necessary ingredients for a theory, over and above the theory's positive assertions, is clarified here.

In Sec. 5 I identify and justify my points of disagreement with Møller-Nielsen's motivational account. In particular, I defend the claim that symmetry-to-reality inferences are sometimes justified even before we have formulated a theory that can be stated only in terms invariant under the symmetry against Møller-Nielsen's objections. Sec. 6 discusses when it is justifiable to posit a symmetry. There remains the concern that the stipulational account cannot do all of the work an account of symmetry must do in physics, since symmetries so defined cannot feature in important theorems. I offer a conciliatory answer to this concern in Sec. 7.

2. Symmetry-to-Reality Inference

Greaves and Wallace (2014: 60) observe and endorse the "widespread consensus that 'two states of affairs related by a symmetry transformation are really just the same state of affairs differently described.'" Although they cite works of philosophy in support of their assertion, one can also find many examples of practicing physicists supporting the same consensus (Dirac 1930; Feynman, Leighton, & Sands 1965: Lec. 17; Strocchi 2008: 119). These statements all amount to assertions of the soundness of symmetry-to-reality inference. To kick off our inquiry into

symmetry, let's unpack what this consensus consists in and what might justify it. This will lead us quickly to some tough puzzles.

To begin with the easiest question, what does it mean in physics to say that two (putative) states of affairs are really the same? Normally we represent a state of affairs in physics with a solution to some equation(s), or a mathematical state in a space of states. An *interpretation* of a theory fixes its content, in the sense of specifying which physical possibility each state represents. This leads naturally to a definition of equivalence: I will say that two states are physically equivalent iff on the correct interpretation of the theory, the states represent the same physical possibility.[2]

What about the other key term, *symmetry*? The definition is often given in colloquial terms: a symmetry is a transformation that preserves a theory's laws. "Preserving" the laws can't just mean something like "If the laws are true of state A, and our symmetry takes A to B, then the laws are true of state B," though. This would mean *every* one-one reshuffling of a theory's states is a symmetry (Belot 2013: 2), and the symmetry-to-reality inference would lead to the absurd consequence that all states are equivalent.

Dasgupta (2016) identifies three families of definitions to flesh out the notion of symmetry. All three require that symmetries preserve the laws in the trivial sense just noted, but they add further conditions:

> *Formal definitions* hold that a symmetry is (defined to be) a transformation on states satisfying some mathematically defined condition.
>
> *Ontic definitions* take a symmetry to be any transformation that preserves some privileged physical properties.
>
> *Epistemic definitions* take a symmetry to be a transformation preserving some privileged epistemically-defined properties (e.g. "the appearances," or the empirical data).

Physicists tend to work with formal definitions. Those texts that don't simply leave the definition ambiguous tend to spell out symmetry in terms of commuting with the dynamical laws: "the action of the symmetry transformation and of time evolution [must] commute" (Strocchi 2008: 7). This is the most plausible type of formal definition, but it gives rise to two very serious problems.

Belot (2013) points out the first problem: the most reasonable ways of spelling out what it is for a symmetry to commute with dynamical evolution

2. I take this to be equivalent in practice to the similar definition given by Belot (2013: 1): "two solutions (models) of a physical theory are *physically equivalent* if and only if, for each possible physical situation, the two are equally well-or ill-suited to represent that situation."

generate counterexamples to the symmetry-to-reality inference. For example, the dynamical laws of classical theories can ordinarily be codified either in terms of a Hamiltonian (energy functional) or a Lagrangian function (encoding the difference between kinetic and potential energy). One can then define a Hamiltonian symmetry as a transformation preserving the Hamiltonian, and a variational (or Lagrangian) symmetry as a transformation preserving the Lagrangian. These amount to two different notions of what it is to commute with the dynamics of a classical theory.[3]

As Belot shows, both definitions give absurd results in certain cases when used in symmetry-to-reality inferences. For example, velocity boosts are not generally a variational symmetry of Lagrangian theories set on Galilean spacetime (Belot 2013: 8). Nor are they a Hamiltonian symmetry, since they alter a system's kinetic energy while leaving its potential energy unchanged (Belot 2013: 12). But the nonexistence of absolute velocity (i.e., the equivalence of states related by Galilean boosts) is a paradigm case of successful symmetry-to-reality inference. And in other cases, both these formal definitions identify as "symmetries" certain transformations that clearly relate physically inequivalent states. For example, *all* solutions of the simple harmonic oscillator are related by a variational symmetry (Belot 2013: 11).

The second problem is that there is no single (known) formal framework in which all physical theories can be written, such that a single formal definition of symmetry can even be specified. As Belot notes, not all classical theories can be described using Lagrangians. Perhaps they can all be given a Hamiltonian formulation. But not all theories are classical theories. Algebraic quantum theories, for example, normally formulate their dynamical laws in terms of automorphisms on an algebra of observables, which do not mention or require either Lagrangians or Hamiltonians (see Summers 2012). Moreover, it is not clear that the notion of time evolution or dynamical laws must apply to all physical theories. It is often suggested that space and time are emergent rather than fundamental quantities in quantum gravity theories (Huggett & Wüthrich 2013). So presumably the fundamental symmetries of these theories should not be defined as transformations that commute with time evolution.

Ontic definitions, on the other hand, are useless for purposes of symmetry-to-reality inference. Dasgupta (2016: 861–66) makes this point persuasively. As he points out, for an ontic definition to be adequate it must completely spell out which physical properties are real. That is, the privileged physical properties included in the ontic definition must at least form a supervenience base for

3. Dasgupta (2016: 864) seems to characterize variational and Hamiltonian symmetries as ontic, but this rests on a restriction of a theory's "formalism" to solely logical vocabulary, which is difficult to defend or apply in physics. Ultimately this is a merely verbal question, but I believe Dasgupta has miscategorized these symmetries.

all of the physically significant properties. Otherwise some symmetries would exist which alter the physically significant properties of a system, and symmetry-to-reality inference would fail for those symmetries. But if we're already in a position to spell out which properties are real, who needs symmetry-to-reality inference? The whole reason symmetry-to-reality reasoning is useful is because we often have no other way to determine which properties are real, aside from checking whether they are invariant under symmetries. So an ontic definition might be extensionally correct, but it will generally be incapable of undergirding the inferential practice we seek to justify—in other words, it cannot really do the work an informative definition of symmetry ought to do.

Dasgupta concludes that only epistemic definitions can succeed, and argues that symmetry-to-reality inference is essentially a special case of Occam's Razor. This is a viable picture of symmetry that I will not attempt to refute here. Instead, I will present a different viable alternative.

3. The Starting Point of Interpretation

My alternative account is best understood within an overall picture of what it is to interpret theories. Here two points will be key. First, interpretation never begins completely from scratch; we always work from what Ruetsche (2011) calls partial interpretations. Second, although the notion of "reality" appealed to in symmetry-to-reality inferences is best spelled out in terms of fundamentality, we are almost always in the business of interpreting non-fundamental theories.

As noted above, an interpretation is a specification of a theory's content, including at minimum a characterization of the physical possibilities represented by its states. It's worth noting that in practice, no interesting theory ever exists in a completely un-interpreted condition. As Ruetsche points out,

> [T]he vast majority of the theories philosophers talk about are already partially interpreted. Otherwise they wouldn't be theories of physics. These theories typically come under philosophical scrutiny already having been equipped, by tradition and lore, with an interpretive core almost universally acknowledged as uncontroversial. (Ruetsche 2011: 7)

For example, it is plausible that for much of its history, the theory of electrodynamics was ambiguous about the ontology of the electromagnetic field. Interpretive work was necessary to determine whether such an entity even exists in any sense. By contrast, it was never ambiguous which aspects of the theory represent charged matter. In its earliest formulations, the theory posited an ontology including electric charge—it came into existence partially interpreted.

Partially-interpreted theories make ontological commitments; they say something about what there is. The early partially-interpreted theory of electromagnetism was ontologically committed to the reality of charge. I see no reason why a partially interpreted theory cannot also say something about what is not real. The early (pre-Minkowski) theory of special relativity provides one example. Plausibly, the theory's commitments included asserting the unreality of absolute rest or absolute simultaneity—this despite the fact that the theory (at least at this early stage) did not include an exhaustive description of its positive ontological commitments. In short, the theory was then only partially interpreted, and among the partial interpretation's commitments was the stipulation that certain putative facts or entities were unreal.

It's natural to understand the partially-interpreted theory of pre-Minkowskian special relativity as stipulating that non-Lorentz-invariant entities and properties are unreal. On the stipulational account of symmetry, this is what it means to say that Lorentz invariance is a symmetry of the theory.

This fits within a rough, idealized account of the average theory's life cycle: First it is proposed as a partially-interpreted formalism. The partial interpretation includes stipulations about which quantities and mathematical structures represent which properties and objects. It may also include stipulations about which quantities and structures do not represent anything real. These are normally formulated in terms of transformations that (the theorists stipulate) preserve all the real structures; in that case, those transformations are called symmetries. Typically the partially-interpreted theory is then empirically confirmed, and once it is sufficiently well confirmed a more complete interpretation is constructed, holding fixed the stipulations made in the partial interpretation. It is at this stage that symmetry-to-reality inferences are useful.

Once a theory's interpretation is complete, symmetry-to-reality inferences are no longer useful, since at that point we have fully determined which of the theory's quantities are real. This need not mean that symmetries have no interpretive purpose at that point, however. To see why they may still be of interest, let's look at the role of fundamentality in interpretation.

On my view, fundamentality needs to come in when we ask the question of what it means for a property or entity to be "real" (as the term is used in symmetry-to-reality inference, and in interpreting physics generally).[4] Here I mean fundamentality in the sense of naturalness or joint-carving (Lewis 1983; Sider 2012). Interpreters of physics are not much concerned with highly non-fundamental facts or things. For example, the question of whether the mereological fusion of two far-apart charged particles that are not in a bound state really exists is

4. One might object that the relevant notion of reality instead boils down to something like objectivity or perspective-independence. Dasgupta (2016: 850–52) argues cogently against this alternative picture.

not a question for interpreters of electromagnetism. Insofar as this question is substantive, it is a question for general metaphysics. The question of whether the electromagnetic field exists, on the other hand, is clearly a question for interpreters of physics. What makes the difference between these two questions? The latter is a fundamental question, because the field (if it exists) is among the most fundamental entities described by electromagnetism. The fusion of two distant particles, on the other hand, is not among the most fundamental entities. Thus its existence is not a substantive interpretive question about electromagnetism.

We must not infer, however, that only perfectly fundamental questions are relevant for interpretation. We know that no currently accepted theory describes fundamental reality in terms of its fundamental properties. If that proves possible, it is a task for a future theory of everything, such as M-theory. In addition, many theories which are less fundamental even than other present-day theories (thermodynamics, classical electrodynamics) are still the subject of interesting interpretive questions. How can we understand this?

The notion of approximate truth is indispensable to scientific realism, and although no satisfactory formal definition exists, our intuitive concept is sharp enough to be useful (Psillos 1999: 261–79). Plausibly, a non-fundamental theory is one which provides a good approximation to the truth in its domain of application, but is such that the truth could be *better* approximated within the same (or a larger) domain by a different (more fundamental) theory.[5] This makes it natural to say that the quantities treated as basic by the non-fundamental theory are approximately fundamental within its domain, in the sense that it is a good approximation to the truth to say that they are perfectly fundamental. So within the domain of thermodynamics (systems with many degrees of freedom), temperature is an approximately fundamental quantity. One does not go very far wrong by treating temperature as one of the basic building blocks when describing such systems. This is what makes temperature foundationally interesting. In general, I suggest that non-fundamental quantities are "real" for interpretive purposes when they are approximately fundamental within the domain of the theory being interpreted.[6]

I promised that these insights about fundamentality would illuminate the interpretive interest of symmetries for a fully-interpreted theory. Here's how: A non-fundamental theory might be formulated using some basic-looking building blocks (primitives) that are not even approximately fundamental. Famously, for example, general relativity is normally formulated in terms of primitive

5. In practice, it may be prohibitively difficult to apply the more fundamental theory in the domain of the less fundamental theory, but all that's required is the possibility in principle of doing so.

6. Some may wish to expand this with the additional principle that quantities definable as functions of the theory's most fundamental primitives (e.g., the square of mass in Newtonian gravity) are also real. I won't commit to that further principle here, but it is compatible with all of this essay's claims.

spacetime points, but it is problematic to treat the identities of these points as real (Hoefer 1996). Facts about the identities of the points are not invariant under the theory's diffeomorphism symmetry, however! So interpreters of the theory should not treat the points' identities as real (approximately fundamental) even though they are among the fundamental-looking constituents of the theory's models (compare Dewar 2015: 321–26). The most complete interpretation—the best statement of what is approximately fundamental in its domain—is given by the theory's models, plus a catalogue of which aspects of the models represent which (approximately) fundamental quantities, *plus* its symmetries. The symmetries do the work of specifying that certain fundamental-looking features of the models are not actually fundamental.[7]

With my account on the table, it is worth noting its close relationship with the picture of symmetry defended by Dewar (2015). On Dewar's view, solutions related by symmetries should be considered equivalent, and "we can implement this [equivalence] without altering our theory, i.e., merely by making acceptable interpretational stipulations regarding the theory" (Dewar 2015: 317). For Dewar, symmetries play the same interpretive role they do on my picture: states related by symmetries are stipulated to be equivalent. But for Dewar, symmetries are not *defined* to be stipulated relations of equivalence. Instead he assumes they can be specified via some mathematical (formal or ontic) definition; he then urges us to stipulate that symmetries so defined relate equivalent states.[8]

---

7. An existing view in the literature which is similarly named but quite different in content is proposed by De Haro and Butterfield (2021), who consider the fact that a theory is often intended by the physicists who formulate it to have certain symmetries—the symmetries are not derived or discovered, but worked into the theory from the start. Consequently they write:

> a theory T [...] is said to have a *stipulated symmetry* if it is formulated as having an automorphism of the state-space [...] that preserves some salient subset of the quantities. The stipulated symmetry thus comes with a choice of which quantities count as salient, so that their values are "worth" preserving.

De Haro and Butterfield are getting at an important distinction about how some symmetries are introduced in scientific practice. But it is clear from the second sentence in their definition that their notion is at most distantly related to the present one, since a symmetry in my sense does not bring with it a choice of some salient quantities that it must preserve.

De Haro and Butterfield's definition would count as an ontic definition, in the terminology I'm borrowing from Dasgupta, because it rests in part on an antecedent list of privileged quantities that are required to be preserved by the stipulated symmetries. Stipulational symmetries in my sense are brute relations of physical equivalence between states, which do the work of telling you *which* of the theory's primitive quantities are real (i.e., relatively fundamental). So on my view a list of the privileged quantities is an "output" that one infers from the symmetries, whereas it's an "input" one employs in defining stipulated symmetries on the Butterfield/De Haro picture.

8. In personal communication, Dewar has confirmed that while he had in mind an ontic or formal definition of symmetry, he is now also sympathetic to stipulational definitions.

Thus, in Dasgupta's terms, Dewar's account rests on an ontic or formal definition of symmetries. This means his account is undermined by Dasgupta and Belot's objections to such definitions, which objections are evaded by my stipulational account.

There is one other significant difference between my stipulational account and Dewar's picture. Dewar (2019) argues that a reduced formulation of a theory, in which symmetries have been eliminated and the theory reformulated in completely invariant terms, may sometimes be inferior to the un-reduced theory for interpretive purposes. For me this is a bridge too far, for reasons that will become clear in the next section.[9]

4. Symmetry and Theories of Everything

Could a fundamental theory of everything have symmetries, as I define them? It depends on how you define 'fundamental theory of everything,' but on the most natural definition the answer is no. I take a theory of everything to be a theory that describes the perfectly fundamental features of reality and nothing else. A theory with (non-trivial) symmetries is not like that. It describes not only fundamental reality, but also some non-fundamental things: the features of the theory's models which are not invariant under the symmetries. The fact that the theory also stipulates (via the symmetries) that these excess features are not fundamental isn't enough to qualify it as a theory of everything. A theory of everything would not mention non-fundamental things at all. To put the point another way, it is nonsense to say that fundamental reality could include brute, unexplained facts about which possibilities are the same or different. But that is what it would mean for a true theory of everything to include symmetries.[10]

But it's entirely possible that the closest humans could ever get to a theory of everything might be a theory with some symmetries. There are a couple of reasons why humans might never be able to formulate a real theory of everything, even in principle. First, it could turn out that there is no such thing as fundamental reality.

9. Another antecedent of the present view is the notion of an "analytic symmetry" as defined by Caulton (2015). But the role of kinematical possibilities isn't central on my account the way it is on Caulton's. So for example, there is room in my account to consider something a symmetry if it holds in all dynamically possible worlds for a theory but fails in some kinematically possible worlds; this would not count as an analytic symmetry for Caulton.

10. To repeat for clarity: this is what it means for a theory of everything to include symmetries *as I've defined them*. It is still possible for transformations defined on a fundamental theory to meet some other definition of 'symmetry.' For example, a fundamental theory could still have *dynamical symmetries*, i.e., transformations that preserve its laws of time evolution. But for the theory to count as completely fundamental, these transformations could not count as symmetries in the stipulational sense of the term—that is, they could not relate physically equivalent states of the theory.

Or relatedly, it could be that there is an infinite descent of more and more fundamental scales which never bottom out in a perfectly fundamental level (Schaffer 2003). These are strange possibilities. (They also don't directly speak to the possibility of symmetries in our ultimate best theory.) What's more likely, in my estimation, is that the true fundamental theory of everything might be ineffable to humans.

I use the term 'ineffable' in the sense discussed by Hofweber (2017): something is ineffable if it is impossible for us to represent it in thought or language. Although he ultimately rejects them for Carnapian reasons, Hofweber presents several cogent arguments for the likely existence of ineffable facts. To my mind, his strongest argument is one of the simplest. Much of what humans know is ineffable to all other animal species on Earth, and even (plausibly) to young human children. What are the odds that human adults, alone out of all known living things, are mentally capable of representing every feature of the universe? It is more plausible that some possible being could do an even better job of representing reality, including its fundamental features, than we are capable of. If this is correct, we should expect there to be some ineffable fundamental states of affairs that could not be represented in any theory entertainable by humans.

This means that the best possible human theory of the universe might describe it redundantly. In particular, it could turn out that two states of such a theory describe the same physical possibility, and the explanation for the equivalence of these states is ineffable to humans. If so, our best possible theory might include symmetries relating such equivalent states.

Could we ever be entitled to posit such a symmetry? It may seem unreasonable to stipulate the equivalence of models in the absence of an explanation. Dasgupta argues that without an invariant reduced theory in front of us, "It remains possible that dispensing with the [non-invariant] feature yields a theory that has too many other vices to warrant belief, such as being too inelegant or complex" (Dasgupta 2016: 854). In such a case, he suggests, we are in the same position as Newton. Because Newton was unaware of the possibility of theories positing absolute acceleration but not absolute position and velocity, he was not in a position to eliminate the latter two quantities from his ontology. By parallel reasoning, one might argue that we can never be justified in positing a symmetry that we cannot explain in terms of a more fundamental invariant theory.

There is room to argue that Dasgupta is mistaken here, and Newton would have been correct to deny the existence of absolute rest (Dewar 2015: 322). Insofar as Newton was justified in inferring that these quantities exist, I would suggest that his justification rested on the assumption (reasonable at the time) that his mechanics was a fundamental theory. If Newton had been alert to the possibility of an unknown or ineffable explanation for the equivalence of different states of absolute rest, on the other hand, he would have had strong reason to posit such equivalence.

In fact, Dasgupta's suggestion that we cannot justifiably posit equivalence in the absence of an invariant theory is implausible. There are cases in which this is clearly reasonable. Think of a mathematician working prior to the development of modern number theory, considering the theory of integers as compared with the integer subset of the rational numbers. Should the mathematician consider the equations "2 + 2 = 4" and "4/2 + 4/2 = 8/2" equivalent in the absence of a satisfactory theory describing their common structure? Surely the answer is yes. Although no "invariant" theory had been formulated at the time, there was every reason to assume its existence.

It can also be reasonable to assume the existence of an ineffable "invariant" theory. Consider a (non-prodigy) seven-year-old child familiar with integers and fractions. A typical seven-year-old brain is unable to entertain all the propositions involved in number theory, so the theory is ineffable to the child.[11] But the child could reasonably come to understand that integer addition and fraction addition must be describing some of the same facts, and might even come to understand that those facts are something the child cannot fully grasp.

The bottom line is that mathematical conjectures, including conjectures about a theory one has not constructed, are sometimes justified. (*How* such conjectures are justified is a tough question, but not one that needs addressing here.)[12] The possibility of such justified conjectures is sufficient to establish the possibility of justified symmetry-to-reality inference even without a reduced theory in hand.

Recent work in the foundations of logic provides a useful (if speculative) example of a place where this sort of reasoning is plausible. As Sider (2012: ch. 10) notes, his realist metaphysics implies that if predicate logic is a fundamental theory, either the universal quantifier ∀ is more fundamental than ∃, or ∃ is more fundamental, or there is redundancy in the world's fundamental logical structure (since every sentence using only ∀ is logically equivalent to some sentence using only ∃ and vice versa).

This is not an attractive consequence of Sider's framework! But McSweeney (2019) presents an alternative option, a thesis she calls Unfamiliar: namely that neither quantifier is fundamental, but that some unknown third structure could ground the truth of both universally and existentially quantified sentences. This "unfamiliar" third structure is clearly something with which we aren't currently acquainted. Indeed, it's plausible that it may be *impossible* for humans to grasp the unfamiliar structure if there is one. After all, our concepts seem to bottom out in existential and universal quantification.

11. Imagine that the child is not part of a community including adults who understand number theory, so there's no sense in which the child can think or speak about complicated number-theoretic facts even by indirect acquaintance.

12. In one of the few systematic treatments of the subject, Corfield (2005: 101–29) suggests that the relevant sort of reasoning is in some ways analogous to Bayesian confirmation.

If this view of the foundations of quantification is correct, this is an example (albeit a non-physical one) in which a symmetry must be posited. And it must be posited on the basis of a conjecture about the existence of an ineffable logical "theory of everything," which would be formulated in terms of McSweeney's unfamiliar structure if only it were possible for humans to entertain it. Thus logic itself may constitute an case where the best humans can do is a theory with symmetries, because the most fundamental theory of its domain is ineffable to us. Ineffability is not a necessary condition for applying this sort of reasoning, of course. It may just be that certain formal tools are comprehensible to humans, but haven't been constructed yet.

A toy example to illustrate this possibility: Suppose that first-order anti- individualist quantifier generalism is the correct picture of meta-ontology (see Sider 2020: 93–105), and so the fundamental truths take the form $\exists x_1$, ...(...x_1...). But no one has yet invented a first-order language with quantifiers. Objects can only be represented using names. Suppose the fundamental truth is $\exists x \exists y(Fx \& Gy)$. Then the closest we could come to expressing the true, fundamental theory would be $Fa \& Gb$, combined with the stipulation that despite appearances, this is equivalent to $Fb \& Ga$.[13]

This example clarifies as well the contribution of symmetries to a theory's content. The reader may have been wondering: if the interpreted formalism of the theory itself includes all of the theory's positive claims about reality, what is added to these claims by stipulational symmetries? Do we really need to say what doesn't exist, in addition to saying what does exist?

What is added is information about which parts of the interpreted formalism's content should not be taken at face value, but should instead be understood as redundant representations of content that we don't (yet?) know how to represent non-redundantly—or that can only be represented non-redundantly in a more fundamental theory.

In the case of the first-order theory without quantifiers we just discussed, the theory has a "symmetry" relating the $Fa \& Gb$ "state" to the $Fb \& Ga$ "state." The theory's formalism by itself implicates that these states are inequivalent; we

13. To avoid contradiction, this stipulation can't be asserted in the same formal language as the sentence itself. So in this case, the equivalent of the best "complete interpretation" (in the philosophy of physics sense, not the model theory sense) of the "$Fa \& Gb$ theory" can only be stated with some use of metalanguage. More generally, theories with stipulational symmetries should be thought of as formulated not just in mathematical language, but in a broader language including both the math *and* sufficiently rich expressive power to stipulate which mathematically inequivalent states are physically equivalent. I take this to be a special case of a broader point made by Maudlin (2018: 6): "*One and the same mathematical apparatus accompanied by a different commentary can convey different physical theories, theories with different ontologies and even with different laws*." (See also Teitel 2021: fn. 26.)

stipulate the symmetry in order to cancel that implicature.[14] So with the symmetry included, the theory conveys different content than it would in the absence of the symmetry.

My position here is similar in some ways to a position Sider (2020) has called the "quotienter" view. The quotienter (a hypothetical character, but one who resembles Dewar among others[15]) holds that "for any model, we can say which features of the model are genuinely representational and which are artifacts. There is no need to provide some privileged, artifact-free description from which we can recover this information" (Sider 2020: 194). Sider contrasts this quotienter with the archetypal fundamentalist metaphysician, who assumes "there must always be some way of describing the phenomenon in question that (in some sense) lacks artifacts. There must be some way of saying what is really going on" (Sider 2020: 194). When I entertain each of these positions, they each have a strong ring of truth to them.

Above I suggested it's possible that the most fundamental theory formulable by humans may contain symmetries. In this sense, I am a quotienter. But I also maintain that if our best theory ends up containing symmetries, the explanation will (probably) have to do with human limitations rather than nature itself. If that's right, there must exist an ineffable theory of everything, which lacks symmetries and hence lacks artifacts—exactly as the fundamentalist metaphysician demands.

The stipulative account of symmetries thus provides a framework for reconciling what seems correct about the quotienter's point of view with what seems correct about the metaphysician's approach. There must indeed be some way, in principle, of saying what is going on—but humans may not be capable of saying it, and hence the best we can do may be to stipulate which features of our best theory are artifacts. (And, to some extent, which of its features are *not* artifacts—e.g., we can perfectly well say that quantities like charge and mass in electromagnetism are real without putting forward a complete reduced theory.[16])

14. Compare Melia (2000), who suggests theorists may employ the trick of asserting something and then "taking it back." I think this is correct, but it also comes at a price: when used in science it can stand in the way of an ideal understanding of the subject matter, as I'll explain in the next section.

15. What Sider calls "quotienting," Dewar (2019) refers to as "sophistication." Dewar's terminology seems preferable here, since "quotienting" is more commonly used to refer to a specific mathematical process for constructing a reduced theory.

16. This also provides a satisfactory answer to the objections Martens and Read (2020) raise against the similar views of Dewar—although it is a concessionary answer, insofar as I grant that Martens and Read are correct when it comes to fundamental theories. Interpreting theories via "sophistication" (quotienting), as Dewar recommends, is an excellent way to garner incomplete information about the approximate metaphysics of a part of reality (namely the domain of a non-fundamental theory). But it cannot succeed as a method of interpreting a theory which is complete, fundamental and exactly true. See also fn. 17 below.

5. Møller-Nielsen's Objection

Some implications of my view have come into question already in the literature on symmetry. In the course of proposing his motivational account of symmetries, Møller-Nielsen (2017) argues against the "interpretational" account on which the existence of a symmetry (formally defined) is sufficient to justify symmetry-to-reality inference. Considering again the example of Newton, he writes,

> The Newtonian who adopts the interpretational construal of symmetries[...] might know that she may legitimately regard all symmetry-related solutions as physically equivalent, but the reality in terms of which this physical equivalence is to be understood will (absent a reformulation of the theory) remain opaque to her; she is offered no immediate explanation as to how such physical equivalence is to be construed or how it could even be said to arise. (Møller-Nielsen 2017: 1263)

All this is equally true of the stipulational account of symmetries. But I don't believe it is a *problem* for the stipulational account, because the consequences Møller-Nielsen points to only seem unacceptable if we imagine them applying to a fundamental theory of everything. In a non-fundamental theory, it is no surprise that some of the interesting (and non-brute) facts posited should be left unexplained. Some such facts may be best explained in terms of a more fundamental theory, and so it may be a mistake to look to the non-fundamental theory for their explanation.[17]

Møller-Nielsen might insist in response that it is unacceptable to posit that two models are equivalent without providing an explanation for their equivalence. This objection could be supported by the claim that we could never be justified in accepting a theory that posits unexplained equivalence.

In response, note first that it may be difficult to make this objection precise without absurdity. Suppose our acceptance of theories is best understood in terms of degrees of belief. Should our degree of confidence in the unexplained equivalence of models always be zero? Presumably not. Is it impossible for any evidence or reasoning whatsoever to bear on the question of whether symmetry-related models are equivalent, in the absence of a known invariant theory?

17. What sort of explanation would the more fundamental theory provide? Møller-Nielsen is on the right track when he alludes to an "explanation [...] as to how such physical equivalence [...] could even be said to arise" (Møller-Nielsen 2017: 1263). A more fundamental reduced theory provides an understanding of how it could possibly be true, consistently and coherently, that the invariant quantities described by the non-fundamental theory are real while its non-invariant quantities are unreal. The non-reduced theory cannot explain this, since it cannot describe the invariant quantities without also saying something about the non-invariant quantities.

Again, this seems a bizarrely overconfident pronouncement for the motivational theorist to make, especially in light of examples like Leibniz's arguments and pre-number theory arithmetic.

To the contrary, it seems that when all we have is a non-fundamental theory, we often have reason to expect that a more fundamental invariant theory will treat certain of the present theory's models as equivalent, even when we don't yet have an invariant theory ready to hand. The grain of truth to the motivational account is that our confidence in the equivalence of symmetry-related states should not exceed our confidence in the existence of an acceptable invariant theory.[18] But that degree of confidence could be quite high even without an invariant theory in hand, if it is based on a plausible conjecture.

6. The Epistemology of Symmetry

All I've done so far is lay out a framework and its consequences. But the framework itself does not tell us when to apply it. Suppose we have a theory T, and are considering positing a symmetry that would identify some of T's states as equivalent, reducing it to T' (which is either known or plausibly conjectured to exist). When should we posit the symmetry?

One way to proceed would be to co-opt the heart of Dasgupta's epistemic account, and posit a symmetry whenever the states related by the putative symmetry transformation are experimentally indistinguishable. This should certainly be a necessary condition for positing a symmetry. But as Dasgupta himself acknowledges, the Occamist reasons he cites for identifying possibilities that aren't detectably different can be outweighed by other theoretical virtues. Thus for Dasgupta, the symmetry-to-reality argument requires "that the hypotheses [states related by the symmetry ...] are equally simple, elegant, common-sensical, and so on; more generally, that they score equally well on every theoretical virtue" (Dasgupta 2021: 6).

Recall Dasgupta's point that without an invariant theory in hand, "It remains possible that dispensing with the [non-invariant] feature yields a theory that has too many other vices to warrant belief, such as being too inelegant or complex" (Dasgupta 2016: 854). I have suggested above that this may be too quick, since we could be in a position to justifiably conjecture that a well-behaved invariant theory exists. But Dasgupta's underlying point is well taken: other virtues matter aside from simplicity. The stipulational account can and should take this underlying point on board.

18. To put this more precisely, in my terms, our confidence in *the existence of a symmetry* should not exceed our confidence in the existence of an invariant theory—since symmetry-related states are equivalent by definition, on the stipulational view.

Thus I conclude that a symmetry (i.e., a stipulation of equivalence between states) should be posited when the overall picture of reality encapsulated by the theory has more and better theoretical virtues with the symmetry than it does without. In the case where the reduced theory (T' above) is known, this boils down to the question of whether T' has more and better virtues than T. Where T' is not known, the question is whether we can justifiably conjecture that T' probably has more and better virtues than T.[19]

Consider again the example of pre-Minkowski relativity. Without making any claims about what Einstein himself believed, it seems to me that the following conclusions were justified at the time: Although no theory had yet been constructed treating the interval as the fundamental quantity, with no mention of position and time coordinates, there was no apparent obstacle to constructing such a theory. Further, its ontology would clearly be more parsimonious than Lorentz's ether, and its laws could be expected to be more parsimonious as well. Thus there was every reason at the time to (tentatively) posit Lorentz invariance as a symmetry in my preferred sense.

I am not, however, convinced by Dasgupta's suggestion that Occamist parsimony will always be the operant theoretical virtue behind symmetry-to-reality inference. Other virtues may be even more important in some cases, for example in the case of gauge symmetry. In electromagnetism, if we were to consider states related by gauge transformations as distinct, the theory would be indeterministic in a pathological-seeming way. The initial state would neither deterministically entail later states, nor would it even entail any probabilistic predictions about later states. This is one of the most important reasons for wanting a gauge-invariant ontology for the theory (Belot 1998: 534–37).

The virtue of a gauge-invariant interpretation is not only parsimony, it is a certain sort of explanatory intelligibility: the initial values of the gauge-invariant quantities can explain the future values of these quantities, while the initial value of the potential is incapable of explaining its future values. This is a significant part of the justification for positing this symmetry.

To give a more speculative example, it is sometimes suggested that some or all of the dualities discovered by the string theory program should be interpreted

19. This means that a point made by Read and Møller-Nielsen (2020), about Dasgupta's epistemic account, applies to my stipulational account as well. Read and Møller-Nielsen point out that there is nothing justifying symmetry-to-reality inference, on Dasgupta's view, other than an application of Occam's Razor. Thus there is nothing indispensable about the concept of "symmetry" in his account; everything it achieves interpretively could be achieved by simply applying the theoretical virtue in question and pointing out which states are empirically equivalent. Similarly, on my own account, symmetry-to-reality inference is simply a special case of applying theoretical virtues that have a broader scope. What the notion of symmetry does, on both my account and Dasgupta's, is to regiment and categorize a family of justified inferences. This makes symmetry a redundant concept, on both my view and his, but I don't see this as a significant objection to either view.

as showing the equivalence of the duality-related states, in a move analogous to symmetry-to-reality inference (De Haro 2019; Huggett 2017). For example, holographic (AdS/CFT) duality has been interpreted, analogously to a symmetry, as showing that string theories in $(D + 1)$ dimensions are equivalent to quantum field theories on their D-dimensional conformal boundary. Plausibly, one virtue of this approach is a sort of unification, both abstract and concrete. At the abstract level, the holographic duality unifies string theories with conformal quantum field theories in a way that is widely considered illuminating. And at the concrete level, it permits an explanation of the otherwise mysterious entropy in blackhole thermodynamics, thereby unifying certain aspects of gravity and thermodynamics.

If correct, this outlook on dualities (construed as symmetries) shows that symmetries can play a role in unifying explanations. Thus the virtue of unification can also be promoted by positing symmetries. This further illustrates that Dasgupta's focus on parsimony is too narrow to capture the breadth of symmetry-to-reality inference as it occurs in science.

Beyond presenting examples like these, I don't believe there is much more that can be achieved in spelling out which theoretical virtues might lie behind symmetry-to-reality inference. There may be a single correct list of the theoretical virtues, and a correct metric of how best to weigh their significance, but we are very far from a systematic understanding of that list.

This is true even though we are rather good at applying the virtues in practice. The situation seems to be roughly analogous to that of normative ethics, where our attempts at systematizing the rules humans ought to live by remain woefully incomplete even though conscientious people tend to be quite accurate at judging which concrete acts are right and wrong except in thorny cases. Many of the most moral people do this without any explicit conscious model of the normative rules they're following. Similarly, good scientists seem to be apt at picking plausible (i.e., virtuous) theories despite lacking an explicit conscious model of the theoretical virtues.

This last point has a further methodological consequence: when scientific experts assert that a certain symmetry-to-reality inference is justified in a given theory, this should be taken as tentative evidence that a symmetry (in my sense) exists. And this is so even when the experts do not cite any explicit theoretical virtues in justifying the inference.

7. Varieties of Symmetry

The stipulational account succeeds, I claim, because it explains why symmetry-to-reality inference is justified, and it has the power to explain this while doing justice to the whole host of reasons that can ground such inference (in contrast to

Dasgupta's purely Occamist account). It also fits in nicely with what I take to be the most plausible account of how fundamental and non-fundamental theories should be interpreted, which account I've outlined above.

But the stipulational account cannot be the whole story about "symmetry" as the term is used in physics. This is because it is manifestly not a formal definition of symmetry, but it is formal definitions that accomplish much of the work done by symmetry in theoretical physics. Consider the famous connection between continuous symmetries and conserved quantities, for example. This connection is established by Noether's theorem, which assumes that the symmetries in question are all symmetries of the action, a formal definition within the Lagrangian framework. Thus only this particular formal definition of symmetry (or a stronger definition that entails it) can do the theoretical work required to explain the nature of conserved quantities. The stipulational definition cannot accomplish this.

This should come as no surprise, however, to anyone who has taken on board the lessons of Belot and Dasgupta's work as summarized in Section 2. For Belot and Dasgupta have shown that no formal definition of symmetry can do the work of grounding symmetry-to-reality inference. Yet, as the example of Noether's theorem shows, formal definitions are required to do much of the work done by concepts of symmetry in interpretively significant areas of physics. Thus no single, univocal concept of symmetry can do all the work that needs to be done. Multiple varieties of symmetry are needed.

There is much that can be said, however, to defuse the threat of heterogeneity that may seem to loom. For it is open to theorists and interpreters of physics to stipulate, wholesale, that certain formal criteria are necessary and/or sufficient for the existence of a symmetry within some given theoretical framework.

In an example that's dear to my own heart, the algebraic approach to quantum theory has its own approach to symmetries (Roberts & Roepstorff 1969). This family of theories represents physical systems using a collection of observables (physical quantities) and a state that assigns probabilities to the observables' different possible values. Dynamical laws are represented by mappings transforming the observables at one time into observables at later times. Symmetries are then understood to be a different set of mappings which permute the observables and which commute with the dynamical mappings (thereby preserving the laws). The stipulational account would take this to be a case of quantum theorists positing that these mappings are relations of physical equivalence because in general, algebraic quantum theories hang together better (are more virtuous) if this stipulation is made wholesale for all the theories within the framework.

And once the stipulation is made, it is possible to prove a quantum version of Noether's theorem (Buchholz, Doplicher, & Longo 1986).[20] Thus (for theories

20. Thanks to Noel Swanson for this example.

within this theorem's domain) it is true, extensionally, that continuous symmetries imply the existence of conserved quantities. This observation is entirely compatible with the stipulational approach to symmetry, and similar examples can be multiplied in other areas of physics.[21]

This example illustrates that, to a significant degree, the stipulational account can co-opt many of the advantages of formal accounts of symmetry, despite the inability of those accounts to ground symmetry-to-reality inference. This makes my view a natural home for those who are tempted by the promise of formal accounts, but who recognize their shortcomings as revealed by Dasgupta and Belot.

8. Conclusions

To zoom out once more, I have argued that symmetry-to-reality inference is best seen as a means for taking a partially interpreted theory and further interpreting it. This is accomplished by stipulating which of the distinctions it seems to draw—which putative possibilities are described by distinct states in its state space—are not real distinctions.[22] Such stipulations should be made on the basis of theoretical virtues. When physicists speak of a theory's symmetries, they are frequently (although not always) making such stipulations.

Interpretive work will be required to prise apart the cases when physicists are using a purely formal or ontic definition of symmetry, as opposed to the present definition. This work is unavoidable, because much of scientific practice rests on symmetry-to-reality inference, but Belot and Dasgupta have shown that no formal or ontic notion of symmetry can justify this practice.

This approach has applications in other areas of philosophy as well. Even many theories in metaphysics that aim at fundamental accounts of reality are only partially interpreted at present. The ultimate truth about these domains may be ineffable, or at least it may not yet have been entertained by theorists. So the positing of symmetries may be essential to progress in these areas of metaphysics as well.

---

21. In a more sweeping (and hence more arguable) example, Wallace (preprint) argues on Occamist grounds that the existence of a dynamical symmetry (as he defines it, a group of transformations commuting with a theory's dynamics) is sufficient to ground symmetry-to-reality inference when the symmetry universally extends from subsystems of the world to measuring devices, or (he suggests more tentatively) when the symmetry is global. If correct, this argument would provide strong grounds for stipulating a symmetry (in my sense) in a very broad framework including all dynamical theories. But Wallace's views rest on a rejection of Belot and Dasgupta's arguments which I would question.

22. That is to say, it is a poor approximation to the truth within the theory's domain of application to say that these states differ in some highly natural way.

The best extant alternative foundation for symmetry-to-reality inference is Dasgupta's epistemic account. My account allows symmetry-to-reality inference to be justified by other virtues rather than just simplicity, which I take to be an advantage over Dasgupta's account. But it's not necessarily a decisive advantage, and Dasgupta may well respond that the difference of opinion here is merely definitional. He could easily grant that other theoretical virtues besides simplicity can also provide reasons for counting states as equivalent; his account simply does not categorize such reasons under the heading of "symmetry considerations."

I have shown, though, that there are more alternatives to ontic and formal accounts beyond just Dasgupta's epistemic picture. Formal accounts, especially, are appealing because of their close relationship to scientific practice, and as we've seen, the stipulational account can co-opt this advantage to a significant extent. The stipulational account has already proven fruitful enough that it deserves serious consideration, alongside epistemic accounts.

Acknowledgements

Thanks to Gordon Belot, John Earman, Michaela McSweeney, Tushar Menon, Bryan Roberts and Laura Ruetsche for detailed and illuminating comments on a previous draft, and to the area editor and two *Ergo* referees for comments that led to significant improvements in the final published version.

References

Belot, Gordon (1998). Understanding Electromagnetism. *British Journal for the Philosophy of Science*, *49*(4), 531–55.

Belot, Gordon (2013). Symmetry and Equivalence. In Robert Batterman (Ed.), *Oxford Handbook of Philosophy of Physics* (318–39). Oxford Univer- sity Press. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780195392043.001.0001/oxfordhb-9780195392043-e-10

Buchholz, Detlev, Sergio Doplicher, and Roberto Longo (1986). On Noether's Theorem in Quantum Field Theory. *Annals of Physics*, *170*(1), 1–17. http://www.sciencedirect.com/science/article/pii/0003491686900862

Caulton, Adam (2015). The Role of Symmetry in the Interpretation of Physical Theories. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *52*, 153–62.

Corfield, David (2005). *Towards a Philosophy of Real Mathematics*. Cambridge University Press.

Dasgupta, Shamik (2016). Symmetry as an Epistemic Notion (Twice Over). *British Journal for the Philosophy of Science*, *67*(3), 837–78.

Dasgupta, Shamik (2021). Symmetry and Superfluous Structure: A Metaphysical Overview. In Eleanor Knox and Alistair Wilson (Eds.), *Routledge Companion to the Philosophy of Physics* (551–62). Routledge.

De Haro, Sebastian (2019). Theoretical Equivalence and Duality. *Synthese*, *198*(6), 5139–77. https://doi.org/10.1007/s11229-019-02394-4

De Haro, Sebastian and Jeremy Butterfield (2021). On Symmetry and Duality. *Synthese*, *198*(4), 2973–3013.

Dewar, Neil (2015). Symmetries and the Philosophy of Language. *Studies in History and Philosophy of Modern Physics*, *52*, 317–27.

Dewar, Neil (2019). Sophistication About Symmetries. *British Journal for the Philosophy of Science*, *70*(2), 485–521.

Dirac, Paul (1930). *The Principles of Quantum Mechanics*. Oxford University Press.

Feynman, Richard Phillips, Robert Benjamin Leighton, and Matthew Sands (1965). *The Feynman Lectures on Physics* (New millennium ed.). Basic Books. https://cds.cern.ch/record/1494701

Greaves, Hilary and David Wallace (2014). Empirical Consequences of Symmetries. *British Journal for the Philosophy of Science*, *65*(1), 59–89.

Hoefer, Carl (1996). The Metaphysics of Spacetime Substantivalism. *Journal of Philosophy*, *93*(1), 5–27.

Hofweber, Thomas (2017). Are There Ineffable Aspects of Reality? *Oxford Studies in Metaphysics*, *10*, 124–70.

Huggett, Nick (2017). Target Space ≠ Space. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *59*, 81–88.

Huggett, Nick and Christian Wüthrich (2013). Emergent Spacetime and Empirical (In) coherence. *Studies in History and Philosophy of Modern Physics*, *44*, 276–85.

Lewis, David (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, *61*(4), 343–77.

Martens, Niels C. M. and James Read (2020). Sophistry About Symmetries? *Synthese*, *199*(1–2), 315–44.

Maudlin, Tim (2018). Ontological Clarity via Canonical Presentation: Electromagnetism and the Aharonov–Bohm Effect. *Entropy*, *20*(6), 465.

McSweeney, Michaela (2019). Following Logical Realism Where It Leads. *Philosophical Studies*, *176*(1), 117–39.

Melia, J. (2000). Weaseling Away the Indispensability Argument. *Mind*, *109*(435), 455–80.

Møller-Nielsen, Thomas (2017). Invariance, Interpretation, and Motivation. *Philosophy of Science*, *84*(5), 1253–64.

Psillos, Stathis (1999). *Scientific Realism: How Science Tracks Truth*. Routledge.

Read, James and Thomas Møller-Nielsen (2020). Redundant Epistemic Symmetries. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *70*, 88–97.

Roberts, John E. and Gert Roepstorff (1969). Some Basic Concepts of Algebraic Quantum Theory. *Communications in Mathematical Physics*, *11*(4), 321–38.

Ruetsche, Laura (2011). *Interpreting Quantum Theories*. Oxford University Press.

Schaffer, Jonathan (2003). Is There a Fundamental Level? *Noûs*, *37*(3), 498–517.

Sider, Theodore (2012). *Writing the Book of the World*. Oxford University Press.

Sider, Theodore (2020). *The Tools of Metaphysics and the Metaphysics of Science*. Oxford University Press.

Strocchi, Franco (2008). *Symmetry Breaking*. Springer.
Summers, Stephen J. (2012). A Perspective on Constructive Quantum Field Theory. https://doi.org/10.48550/arXiv.1203.3991
Teitel, Trevor (2021). What Theoretical Equivalence Could Not Be. *Philosophical Studies*, *178*, 4119–49.
Wallace, David (preprint). "Observability, Redundancy and Modality for Dynamical Symmetry Transformations".