Forthcoming in *Synthese*

Are Models Our Tools Not Our Masters?

Caspar Jacobs*

Abstract

It is often claimed that one can avoid the kind of underdetermination that is a typical consequence of symmetries in physics by stipulating that symmetry-related models represent the same state of affairs (Leibniz Equivalence). But recent commentators (Dasgupta 2011; Pooley 2021; Pooley and Read 2021; Teitel 2021a) have responded that claims about the representational capacities of models are irrelevant to the issue of underdetermination, which concerns possible worlds themselves. In this paper I distinguish two versions of this objection: (1) that a theory's formalism does not (fully) determine the space of physical possibilities, and (2) that the relevant notion of possibility is not physical possibility. I offer a refutation of each.

1 Introduction

It is often thought that the presence of symmetries in physical theories entails an undesirable form of underdetermination, as well as, in certain cases, indeterminism. In brief, let the *space of models* of a theory consist of mathematical structures that are used to represent ways the world could have been if the theory were true.[1] The *symmetries* of a theory are certain transformations that preserve the space of models: if M and M′ are related by a symmetry, then M is a model of the theory iff M′ is.[2] In many cases, symmetry-related models are empirically equivalent. This is the case for the spacetime symmetries of both classical mechanics and general relativity, as well as for gauge symmetries. Consequently, theories that display symmetries seem to exhibit a form of underdetermination: for each physically possible world, there exist distinct yet empirically equivalent physically possible worlds (Fig. 1).

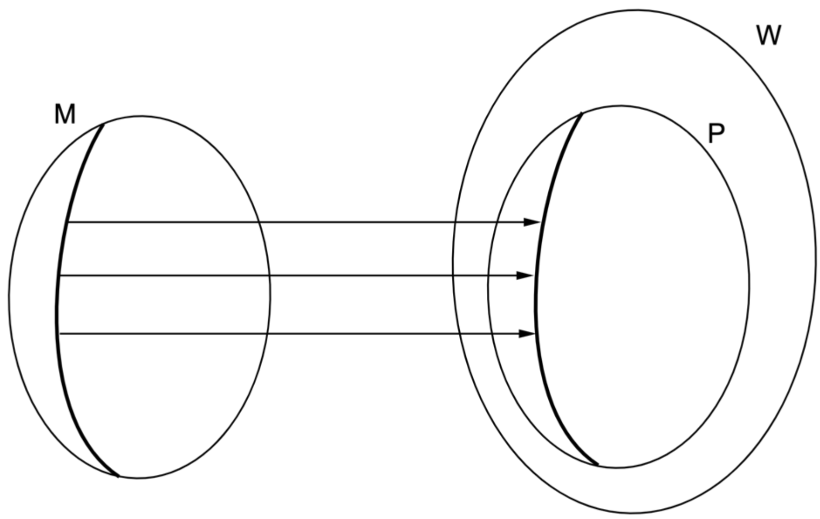

Figure 1. M *represents a theory's space of models;* W *represents the set of possible worlds, and* P *the subset of physically possible worlds. The curved line within* M *represents an* orbit *of symmetry-related models, and the curved line within* P

* Institute for Philosophy, Leiden University, Nonnensteeg 1-3, 2311 BE, Leiden, The Netherlands

[1] I here adopt what Wallace (2019a) calls the 'Cosmological Assumption': that a theory's models represent states of the entire universe. A similar problem occurs for Wallace's *subsystem-local* symmetries, so one could rephrase the argument of this paper in terms of them instead.

[2] This is only a necessary condition on symmetries. To develop a complete definition, one would have to specify *which* transformations are relevant. This lies beyond the scope of this paper; for two very different attempts, see Dasgupta (2016) and Wallace (2019b).

likewise represents an orbit *of empirically equivalent worlds. When the map from models to worlds is one-to-one, underdetermination ensues.*

The standard solution to this problem is to endorse a principle known as Leibniz Equivalence.[3] Firstly, define a *representational convention* as a function from the theory's space of models to the possible worlds. The range of this map is the set of *physically* possible worlds. Leibniz Equivalence is then formulated as follows:

> **Leibniz Equivalence**. If models M and M′ are symmetry-related, then under any representational convention C they represent the same possible world W.

If M and M′ merely represent "the same state of affairs differently described" (Greaves and Wallace 2014), then the presence of symmetries does not entail the existence of empirically equivalent physically possible worlds: underdetermination is averted (Fig. 2).

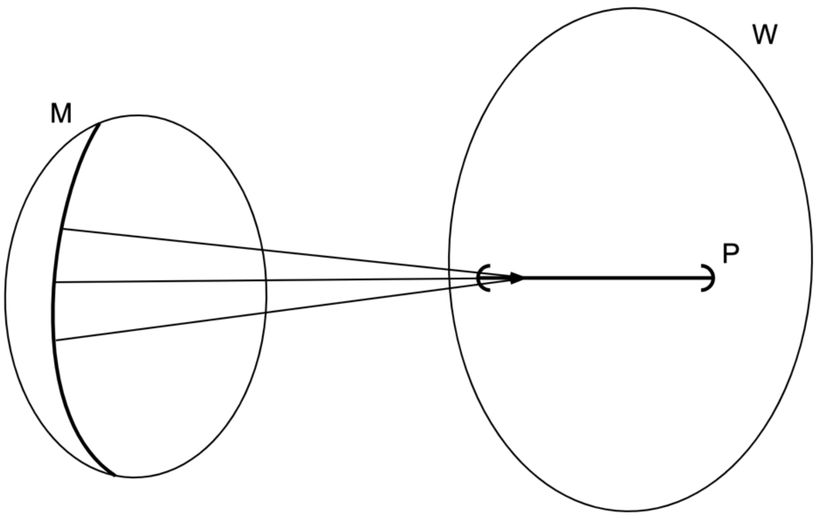

Figure 2. Here, the map from models to worlds is many-to-one: models within an orbit represent the same world from a 'reduced' space of physical possibilities. There is no underdetermination.

Leibniz Equivalence is not intended as an *a priori* claim about the representational capacities of symmetry-related models. Many advocates of Leibniz Equivalence admit that such models *could* represent distinct worlds. Rather, Leibniz Equivalence amounts to the *stipulation* that one will use symmetry-related models to represent the same possible world.[4]

That is how the story is usually told. But these days, it is fashionable to deny that talk of models is relevant to the problem of underdetermination (Dasgupta 2011; Pooley 2021; Pooley and Read 2021, Teitel 2021a). The objection is that Leibniz Equivalence is merely a thesis about the way we use models to represent possibilities, not one about possibilities themselves. On the other hand, the underdetermination problem *does* concern possibilities. So, theses such as Leibniz Equivalence cannot bear on underdetermination. We are free to use a theory's models however we please: they could represent physical possibilities, but also the streets of Paris. In Dasgupta's (2011, 134) rousing words: "models are our tools not our masters".

[3] Leibniz Equivalence is professed, in one form or another, by Saunders (2003), Baker (2010), and Greaves and Wallace (2014). For recent discussions of the principle, see Roberts (2020) and Jacobs (2021). Notice that Roberts' definition of Leibniz Equivalence is different from mine, since it concerns *isomorphic* models. These definitions coincide when a theory's symmetries *are* isomorphisms, as is the case for general relativity. There is a distinct literature on the representational capacities of isomorphic models: cf. Weatherall (2018), Fletcher (2020) and Pooley and Read (2021).
[4] I will address the question of whether such a stipulation is always warranted in Section 4.

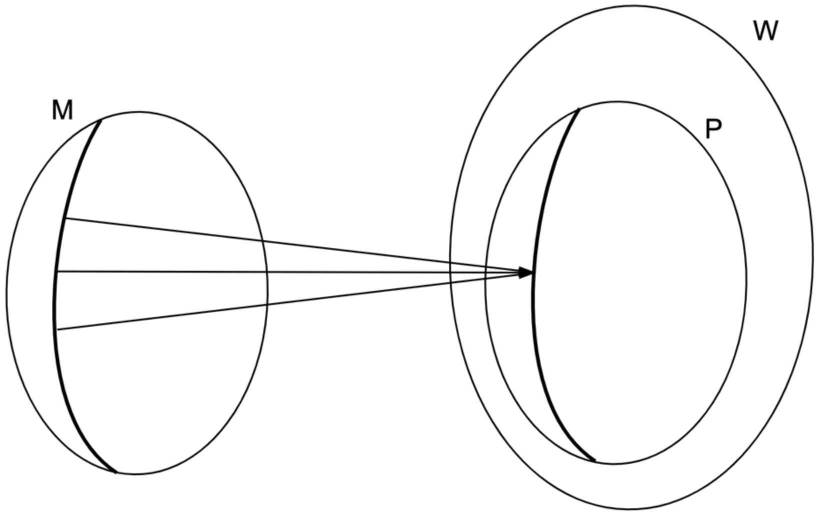

Figure 3. As before, the models on an orbit represent the same physically possible world. But in addition, there are empirically equivalent physically possible worlds not represented by any of the theory's models: underdetermination looms.

This objection can be sharpened in two ways. On the one hand, the point may be that the formalism of a theory tells us little about what is physically possible. This can happen if one does not believe that each physically possible world is represented by some model of the theory. It then follows that Leibniz Equivalence does not constrain the space of physically possible worlds: the existence of empirically equivalent (yet numerically distinct) physically possible worlds is consistent with Leibniz Equivalence (Fig. 3). For example, Leibniz Equivalence implies that boost-related models represent the same physically possible world. But this does not rule out that those models all represent a world in which the universe is at absolute rest, while there is another physically possible world in which the centre of mass is in motion not represented by *any* of the theory's models. Since absolute velocity remains unobservable in this scenario, Leibniz Equivalence has not prevented underdetermination.

Both Dasgupta (2011) and Pooley (2021) can be read as making this point:

> For a substantivalist might grant that there are distinct [symmetry-related] worlds W and d(W) but simply deny that they are represented by [symmetry-related models] M and d(M) respectively. (Dasgupta 2011, 126)

> Premise (1) [that symmetry-related models represent distinct possibilities] is, therefore, best thought of as the combination of two theses: one about the plurality of possibilities [...]; and another about how particular mathematical objects represent those possibilities. Ultimately, it is only the first thesis that does essential work [...]. (Pooley 2021, 149)

Dasgupta and Pooley seem to suggest that a theory's models need not represent all of the theory's physically possible worlds. It follows that a constraint on the representational capacities of models need entail a restriction on the space of physically possible worlds. Teitel's (2021a) objection, discussed below, also has this flavour. These authors seem to hold a relatively loose view of the connection between models and possibilities, on which facts about the latter are significantly independent from considerations of the former. In Section 2, I will argue that this view is mistaken.

On the other hand, it is unclear from the context whether 'possibility' here refers to *physical* possibility or some broader notion—call it *metaphysical possibility*. Read in this latter way, one may concede that one can read off physical possibility from a theory's formalism, yet deny that it is this form of possibility that is relevant to the underdetermination problem. On this view there is a threat of underdetermination as soon as there are symmetry-related possible worlds, *even if some of those worlds are unphysical* (Fig. 4).[5] This is perhaps better understood as a case of underdetermination of theory by the data, rather than one of the possible world within a theory. Nevertheless, the former may seem equally worrisome as the latter. In Section 3, I will argue that it is not.

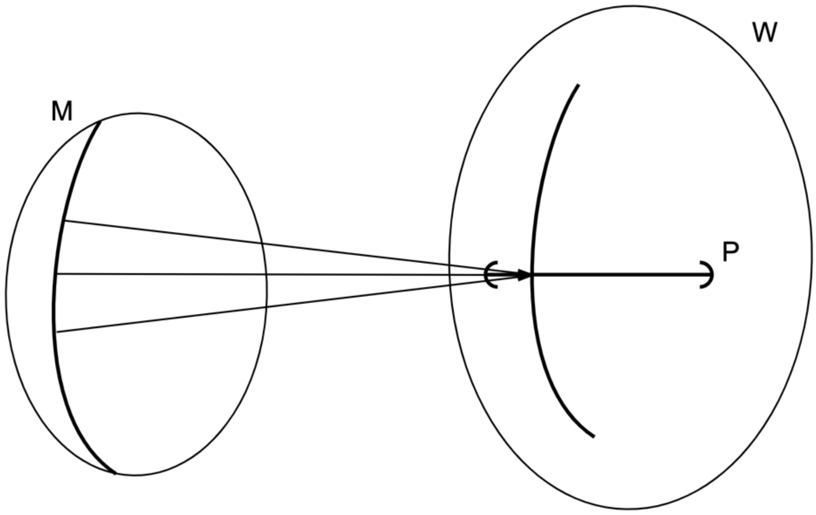

Figure 4. The models within an orbit represent a unique physically possible world; there are no empirically equivalent physically possible worlds. But there are *empirically equivalent possible worlds that are not physical. Is this a case of underdetermination?*

Finally, in Section 4, I discuss the broader lessons for the metaphysical interpretation of a theory's symmetry-related models, namely that Leibniz Equivalence imposes a restriction on the metaphysical picture they present.

2 Interpretation and Representation

There is one sense in which it is true that a theory's models are wholly uninformative of physical possibility, namely that they need not represent physical possibilities at all. While it is perhaps not fully true that we can use any model to represent "just about any physical situation whatsoever" (Teitel 2021b, 4137), Putnam's infamous paradox suggests that any model may represent any possibility up to cardinality. The models of classical mechanics, for instance, can be used to represent worlds in which classical mechanics is true, but also the history of the Dutch Republic 1588-1795. The consequence is that a thesis such as Leibniz Equivalence cannot constrain the space of physically possible worlds. Leibniz Equivalence says that symmetry-related models represent the same state of affairs, but if they are used to represent Dutch history this has little to bear on what is physically possible.

However, this sense in which models are uninformative is also quite anodyne. The interpretation of physical theories does not proceed in isolation, but against the backdrop of certain plausible principles. It is perhaps a philosophically worthwhile exercise to consider the

[5] Isn't it definitional of symmetries that they preserve satisfaction of the laws? Not necessarily: Dasgupta (2021) discusses so-called 'empirical symmetries' that relate worlds with distinct yet empirically equivalent laws. For the sake of argument I am happy to accept this broader notion of symmetries.

representational capacities of models stripped from any context. For example, Teitel (2021a, 9) argues that on certain representational conventions (we will see an example below), Leibniz Equivalence is moot: "these representational conventions may not be our own, but their coherence is all we need to bring out the conceptual rift between [Leibniz Equivalence] and the target desideratum of no-shifts itself". But the *internal* coherence of these conventions is insufficient. If it is the interpretation of physical theories that interests us, we should consider representational conventions that are consistent with the practice of interpretation. If they are not, then their existence is unlikely to move interpreters of physical theories.

The question thus becomes: *given certain reasonable constraints on theory interpretation*, does Leibniz Equivalence entail that there are no empirically equivalent physically possible worlds? The remainder of this section defends an affirmative answer to that question.

2.1 The Standard Account of Interpretation

To answer the question we have to know the relevant constraints. Consider what Ruetsche (2011, 7) calls the 'Standard Account' of interpretation: "to interpret a theory is to characterize the worlds possible according to it". The idea, Ruetsche explains, is that one can identify a proposition with the set of possible worlds in which it is true. The content of a theory then consists of the possible worlds in which the theory's propositions—the laws—are true. But the Standard Account is not wedded to this account of propositions; it only requires the weaker thought that to understand a theory is to know "what the world would be like if the theory were true" (Earman 2004, 1234). I will assume the Standard Account here.[6] This account constrains the practice of theory interpretation because it provides a particular account of what interpretation *is*.

In more detail, I take the following pair of principles to capture the Standard Account:

> **Soundness**. For any reasonable representational convention C: if M is a model of theory T, then M represents a world W that is physically possible according to T under C.
>
> **Completeness**. For any reasonable representational convention C: if W is a world that is physically possible according to T, then there is a model M of T that represents W under C.

These principles implicitly define what it is for a representational convention to count as 'reasonable'.

Soundness and Completeness allow one to infer physical possibilities from theoretical models and vice versa. Soundness simply says that the models of a theory are intended to represent physically possible worlds, rather than, say, Dutch history. This is just what it means to interpret a theory's formalism *as* a physical theory. Completeness says that any physically possible world is indeed represented by one of the theory's models. If this weren't the case, the interpretation would be incomplete: it would not specify the theory's full content (by some prior standard of what that content is, such as consistency with experimental evidence).

[6] The Standard Account has recently come under attack from Ruetsche (2011), Williams (2019) and Wallace (2021a). Ruetsche believes that interpretation should take into account 'pragmatic' factors; Williams argues that it should take into account inter-theoretic relations; and Wallace claims that theories chiefly represent subsystems of the universe. Yet these criticisms are consistent with the core idea that to interpret a theory is to characterise a set of possible worlds (or perhaps, in the case of Wallace, *parts* of possible worlds). Indeed, Ruetsche explicitly states that she is "not ready to give up" on this feature of the Standard Account. For an articulate defence of the Standard Account, see Belot (1998, §6).

Soundness is relatively uncontroversial, and in any case it doesn't bear on the present debate.[7] But Completeness does, as I will discuss now.

2.2 Completeness & Leibniz Equivalence → No Underdetermination

Given Completeness, Leibniz Equivalence entails that there are no empirically equivalent physically possible worlds (the situation in Fig. 2 rather than that in Fig. 3 obtains). First, fix a representational convention. The argument proceeds under this convention. Then:

1. If W and W′ are physically possible worlds, then there is a model M that represents W and there is a model M′ that represents W′. (Completeness)
2. If M represents W and M′ represents W′, then if W and W′ are empirically equivalent then M and M′ are symmetry-related.[8]
3. If M represents W and M′ represents W′, then if M and M′ are symmetry-related then W=W′. (Leibniz Equivalence)
4. Therefore, if W and W′ are empirically equivalent, then W=W′.

The argument is valid; and I have argued that Completeness is a reasonable interpretative principle. Therefore, Leibniz Equivalence entails that there is no problem of underdetermination.

What has led the authors quoted in the introduction to reject the efficacy of Leibniz Equivalence? It is now clear that Dasgupta's and Pooley's claims violate Completeness. Both believe that it is possible for a theory's models not to represent *all* physically possible worlds. If that is the case, the argument fails: it would only establish the weaker conclusion that *of the worlds represented by the theory's models* none are empirically equivalent. But if there are more physically possible worlds than that, underdetermination still looms.

The rejection of Completeness is also explicit in an argument from Pooley and Read:[9]

> Suppose that C is a class of [empirically equivalent] worlds—and make no assumptions […] about the cardinality of C, which might perhaps contain only a single member. Let M and M' both be members of a class of [symmetry-related] spacetime models of a type and particular character that makes them apt to represent members of C. Someone who embraces [Leibniz Equivalence] […] disavows being able to use M and M' to jointly describe different members of C (if different members there are). But they do not thereby save [the theory in question] from [underdetermination]. If C really does contain a plurality of members […], then [the theory] […] does not distinguish between the possibilities even to the extent of not being able to refer differentially to them. It therefore (implicitly) regards them as all equally possible, which is just to say that […] the theory is [underdetermined]. (Pooley and Read 2021, 24)

Pooley and Read in effect assume that we can separately consider a theory's physically possible worlds, C, and the class of worlds represented by the theory's models. Their argument then proceeds from the claim that the former can outstrip the latter, so that Leibniz Equivalence has

[7] Soundness is not universally accepted. In particular, approaches that accept Butterfield's (1989) '(One)', such as Maudlin's (1988) metric essentialism, are incompatible with Soundness.

[8] This assumes that only symmetry-related worlds are empirically equivalent; but one can restrict this premise and the conclusion to just those empirically equivalent worlds that are related in the relevant sort of way, for instance by a Galilean transformation.

[9] The scope of Pooley and Read (2021) is slightly narrower, as they only consider diffeomorphic models of general relativity. I have amended the quote to apply more generally. Even if Pooley and Read would not advocate these more general claims, however, they still reject Completeness in their treatment of general relativity in particular.

no effect on which worlds are members of C. This claim clearly violates the Standard Account, since Completeness entails that if the models of a theory don't represent some world in C, then C is not the class of physically possible worlds.

Insofar as Completeness is one of the reasonable principles that guide the act of theory interpretation, then, Pooley and Read's argument fails. The same is the case for the claims from Dasgupta and Pooley quoted in the introduction. Neither Dasgupta, nor Pooley or Read address Completeness in any way. I have argued that Completeness is part and parcel of the Standard Account of theory interpretation, so bar any reason to depart from the standard on this particular point the conclusion remains that Leibniz Equivalence does entail the absence of empirically equivalent physical possibilities.

I will now discuss two ways in which one could avoid this conclusion. Both are ultimately unsatisfactory.

2.2.1 Reject Completeness?

The easiest way out is to reject Completeness. In particular, one could argue that the universal quantification over representational conventions in Completeness is too strong, and accordingly replace it with an existential quantification. This yields:

> **Weak Completeness**. If W is a world that is physically possible according to T, then there is a model M of T that represents W under some representational convention C.

According to Weak Completeness, it is possible that under no representational convention the models of a theory collectively represent all physically possible worlds. The full space of physically possible worlds instead consists of the worlds that are represented by any of the theory's models under *some* representational convention. Perhaps under one convention an equivalence class of symmetry-related models represents a certain world, and under a different convention it represents a distinct yet empirically equivalent possible world (Fig. 5).

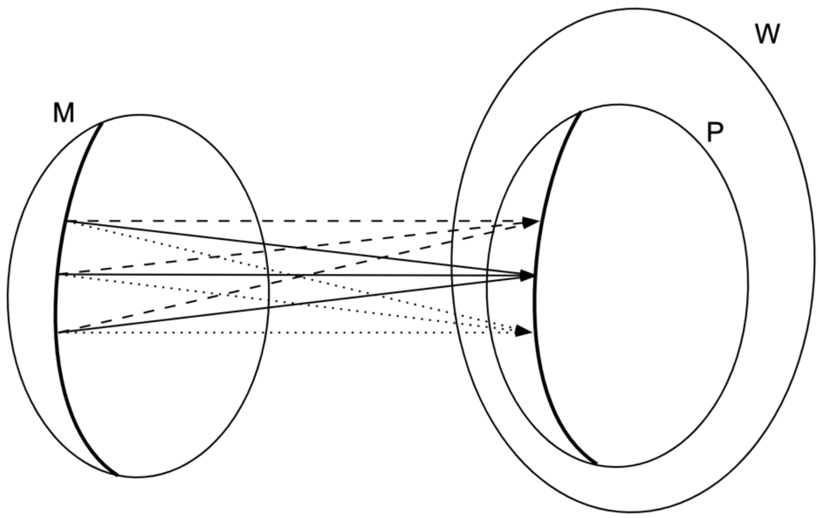

Figure 5. The arrows of different types (unbroken, dashed, dotted) indicate different representational conventions. On each convention, models within an orbit represent the same possible world; but this world is different on different conventions.

With Completeness so weakened, underdetermination looms once more. Although no pair of symmetry-related models represents distinct yet empirically equivalent possible worlds under the same representational convention, the same equivalence class of symmetry-related models may represent distinct yet empirically equivalent possible worlds under different representational conventions. Since Weak Completeness only demands that physically possible worlds are represented by the theory's models under *some* representational convention, all of those worlds are physical. Leibniz Equivalence is nevertheless satisfied, hence it does not prevent underdetermination.

But Weak Completeness is a questionable principle to endorse. It implies that there is no way to fix the referents of the theory's names, predicates and relations such that the theory can represent all physically possible worlds 'at once'. It is impossible to simultaneously represent a pair of symmetry-related worlds without changing the meaning of the theory's terms halfway through. Put differently, there is no way to describe the difference between these worlds within the theory's language; there is no sentence of the theory which, *keeping the interpretation fixed*, is true of one world but false of the other.[10] In such a situation, it is more natural to expand the theory's expressive resources to include structure that distinguishes between such worlds. Once that structure is added, however, the models in question are not symmetry-related anymore, since the additional structure is (by construction) not preserved across them. Therefore, no symmetry-induced problem of underdetermination arises in this case either.

I emphasise that it is not my aim to prove that Completeness is absolutely unassailable. I only defend the weaker claim that Completeness is a reasonable principle in the interpretation of physical theories, whereas Weak Completeness is not. This is borne out by the fact that, as far as I am aware, no one has ever advocated a principle like Weak Completeness. This is sufficient for Leibniz Equivalence to solve the problem of underdetermination.

2.2.2 Reject Uniqueness?

There is another, perhaps less radical way to avoid the conclusion that Leibniz Equivalence constrains the space of physically possible worlds: reject the assumption, implicit in Soundness and Completeness, that a model represents only *one* world under any representational convention. When that assumption fails, we cannot refer to *the* world represented by a model. This means that Leibniz Equivalence cannot prevent underdetermination. For even if the models within an equivalence class of models [M] closed under symmetries have the same representational capacities, they may *each* represent a variety of empirically equivalent possible worlds (Fig. 6).

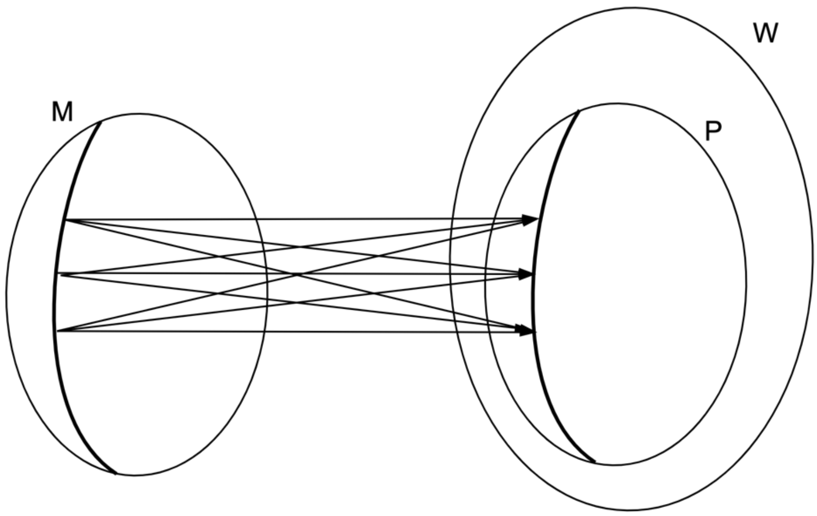

Figure 6. The models within an orbit still represent the same possibilities, but the representation-relation is many-to-many.

[10] This claim echoes the conclusions of Bradley and Weatherall (2020).

This is the gist of Teitel's (2021a) objection:[11]

> Notice that there is nothing incoherent about [someone] who accepts [Leibniz Equivalence] yet rejects [that there are no symmetry-related possibilities]. To do so, she need only adopt representational conventions governing how to use the formalism of [the theory] that predict [Leibniz Equivalence], and there are many options for doing so. For instance, perhaps she thinks that, although there are [symmetry]-related nomic possibilities, for every equivalence class of [symmetry]-related mathematical solutions we use each member of the equivalence class to represent the same qualitative proposition […], namely the qualitative proposition which is true at all and only the members of the corresponding equivalence class of [symmetry]-related nomic possibilities. (Teitel 2021a, 9)

Here, Teitel's nomic possibilities are equivalent to our physical possibilities.

Of course, it is uncontroversial that the same model can represent different states under distinct representational conventions (Fletcher 2020). Teitel's objection turns on the claim that one model can represent distinct possible worlds even under the *same* representational convention. This happens when a model does not represent all physical features of the world. For example, consider a model of classical mechanics set on Galilean spacetime. On a standard representational convention, such a model represents particles with absolute accelerations but without absolute velocities. But on a different representational convention the same model may represent an equivalence class of possible worlds related to each other by uniform boosts, that is, possible worlds in which particles have different absolute velocities. On this view the theory's models simply leave out a standard of absolute rest, which is nevertheless considered as real physical structure.

Teitel's counterexample is consistent with Soundness and Completeness. But it does violate another principle:

> **Uniqueness**. If M is a model of theory T, then under any representational convention it represents a unique physically possible world W.

For Uniqueness to fail would mean for the world to contain features that are not represented in the theory's models; that even a full interpretation of a model's structure leaves the possibilities it represents partially unspecified.

Putting Soundness, Completeness and Uniqueness together, it follows that: under any representational convention, each model represents a unique physically possible world; and each physically possible world is represented by some model. There is thus a surjective function from the space of models onto the physically possible worlds. This leaves open whether that function is injective or not, that is, whether models represent worlds one-to-one or many-to-one. In either case, however, it is impossible for one model to represent distinct possibilities.[12]

Is Uniqueness plausible? It is difficult to say, because the principle is so widely-accepted that explicit discussions are all but absent from the literature. The only exception of which I am

[11] Teitel considers only the diffeomorphism symmetries of general relativity. I have amended the quote to cover symmetries more generally. Even if Teitel would not advocate this more general claim, however, he still rejects Uniqueness in his treatment of general relativity in particular.

[12] Uniqueness requires some modification if it is to apply to subsystem states, since the same model can represent different subsystem states when coupled to the environment. For models of subsystems, then, the relevant principle is that each model represents a unique state *when coupled to a fixed environment*.

aware is Butterfield (1989), who argues that Uniqueness is entailed by physicalism: the only way in which the same model could represent distinct possibilities is if there is more to the world than just physics. It doesn't seem to occur to Butterfield that the same model could also represent distinct possibilities that differ in their physical features. Moreover, what if physicalism were false? Would the presence of symmetries then entail the underdetermination of a theory such as general relativity? It seems odd to suggest that the truth of physicalism could have any effect on such mundane matters. For the same reason, it seems unnecessary to endorse physicalism just to avoid symmetry-induced underdetermination.

Once one realises that anti-physicalism could derail Uniqueness, one can spot many other forms of 'underdetermination' consistent with our theories. Perhaps the thoughts of the Gods are underdetermined by the physical facts; or perhaps the movements of invisible ghosts are such that one model may represent many; or perhaps there even are some physical fields completely decoupled from any empirically accessible quantities. The point is convincingly made by Norton (2020), who puts it in the form of a *reductio*: if so-called symmetry-to-unreality inferences rule out the existence of, say, spacetime points—as alleged by Earman and Norton (1987)—then, absurdly, they must also rule out the existence of Gods, ghosts and any other entities not explicitly represented within the formalism of our physical theories.

These forms of underdetermination need not worry us. Following Norton (2020), the kind of underdetermination one should care about is *internal* underdetermination—underdetermination of the quantities internal to the theory's dynamics—and however one defines that term of art it should surely exclude the above examples. To get a better grip on the difference, note that models of Newtonian mechanics set on Galilean spacetime can only represent mental states or states of absolute motion in an unusual way: *in absentio*. Sticking with the latter example, there is no element of the theory's models that one can point to and say: "*that* represents the particle's state of absolute rest". The same is the case for the non-qualitative propositions of Teitel's scenario. If one is already committed to the claim that, under any representational convention, symmetry-related models represent the same class of possible worlds, then there cannot be any element of the theory's formalism that represents haecceities—even if that very class of models is used to represent haecceitistically distinct worlds. The models are simply silent about them. Call the objects, quantities, and structures that are explicitly represented by our theory's models *theory-internal.*[13] Then the kind of underdetermination that is relevant, internal underdetermination, occurs only when the values of the theory-*internal* quantities are underdetermined by the empirical facts.

Furthermore, call possible worlds *internally equivalent* whenever they are equivalent insofar as their theory-internal quantities are concerned. It is an immediate consequence that:

> **Internal Uniqueness**. If M is a model of theory T, then under any representational convention it represents a unique equivalence class [W] of physically possible worlds up to internal equivalence.

Internal Uniqueness is consistent with Teitel's counterexample, as well as with the possibility that physicalism is false. For in either case, the differences between the possibilities represented by one and the same model are non-internal.

Nevertheless, Internal Uniqueness suffices to avoid any worrisome form of underdetermination. For it is underdetermination of the theory-internal quantities that one should care about;

[13] It follows that the set of theory-internal quantities is also theory-relative: microscopic degrees of freedom are internal to kinetic theory but external to thermodynamics.

underdetermination of those facts that our physical theories are *about*. The problem of underdetermination is that there are *internally* distinct yet empirically equivalent physically possible worlds, that is, physically possible worlds that are empirically alike yet differ over the values of one of the quantities explicitly represented by the theory. But Soundness, Completeness, and Internal Uniqueness jointly entail that if Leibniz Equivalence is true, then symmetry-related physically possible worlds are *identical up to internal equivalence*. It follows that the theory is not underdetermined insofar as the facts within the theory's domain of discourse are concerned. Therefore, these three reasonable principles of theory interpretation jointly entail that an endorsement of Leibniz Equivalence suffices to solve the only problem of underdetermination worth the name.

To finish this section, note the parallel between my response to Teitel's objection and certain responses to the Hole Argument that advocate qualitative definitions of determinism. Recall that in the Hole Argument, the facts at some time *t* do not uniquely fix the facts thereafter: indeterminism ensues. The facts left unfixed are haecceitistic ones. Brighouse (1997) and Melia (1999), amongst others, have argued that the relevant definition of determinism should only concern 'physical' facts—and that haecceitistic facts are not physical. This is parallel to our restriction to internal quantities. This response to the Hole Argument has struck many as unsatisfactory (Belot 1995, Brighouse 2020). In particular, the claim that haecceitistic facts are unphysical seems arbitrary, designed just to avoid the spectre of indeterminism. The restriction to internal quantities, on the other hand, is not arbitrary. It follows from the idea that the determinism of a theory should only concern whatever that theory's models explicitly represent, that is, whatever that theory is *about*. Notice, for instance, that my criterion does not entail that haecceitistic facts are always external. They are if one embraces applies Leibniz Equivalence to the diffeomorphism-related models of GR, since then such models at most represent qualitative facts. But if one rejects Leibniz Equivalence and hence allows for the possibility that symmetry-related models represent haecceitistically distinct possible worlds then haecceitism *does* contribute towards the theory's internal indeterminism, since in that case the points of the spacetime manifold are taken to explicitly represent haecceities. This is not a problem for my approach, however, since it is exactly the claim that Leibniz Equivalence enables a solution to issues such as indeterminism that I wish to defend.

3 Possibility and Detectability

In the previous section I showed (i) that Soundness, Completeness and (Internal) Uniqueness are reasonable principles of theory interpretation, and (ii) that conditional on those principles, Leibniz Equivalence entails that symmetry-related models represent classes of internally equivalent physically possible worlds. This suffices to avoid the relevant kind of underdetermination, namely underdetermination of theory-internal quantities by the empirical data.

The conclusion is premised on the claim that only *physically* possible worlds 'count' for underdetermination. The second version of the objection from the introduction states that one faces a problem of underdetermination as soon as there are pairs of empirically equivalent yet internally distinct possible worlds—whether *physically* possible or not. Consider, for instance, a theory T which adds to classical mechanics the postulate that the centre of mass of the universe is at absolute rest. Suppose that W is a physically possible world of T, and that W′ is another possible world just like W except that the velocities of all bodies in W′ are boosted. Because absolute velocity is a quantity internal to T—the laws of T are partially *about* absolute velocity—these worlds are internally distinct. Yet W′ is not a physical possibility of T: the universe's centre of mass moves! The second objection claims that the possibility of such a world nevertheless poses a threat of underdetermination. If so, Leibniz Equivalence does not preclude underdetermination after all, for the previous section established only that Leibniz Equivalence entails that there are no empirically equivalent yet internally distinct *physically* possible worlds. To be sure, this is a different kind of underdetermination. It is not the choice between symmetry-related models of the same theory that is underdetermined by the data, but one between systematically related models of *different* theories. The aim of this section is to show that this is not a worrisome form of underdetermination.

The example from the previous paragraph is a little reserché: there is no law that says that the centre of mass has a certain absolute velocity. However, there are similar but more realistic cases. Maudlin's (1988) metric essentialism, for instance, entails that in general relativity it is *physically impossible* to simultaneously translate the matter and the metric field, despite the fact that such a transformation is 'allowed' by the theory's equations. Yet a shifted world is *metaphysically possible* on Maudlin's view. So, for Maudlin there are empirically equivalent yet internally distinct possible worlds; it's just that not all of those worlds are physical. Furthermore, from the broader literature on spacetime symmetries it would seem that the chief concern is the mere existence of symmetry-related possibilities whether physical or not: doctrines such as relationism or anti-haecceitism do not just entail that boosts are forbidden by the laws of physics, but that such transformations are simply impossible. If the actual world fundamentally consists of spatiotemporal relations between bodies, then there just *are* no absolute velocities to boost. There is no world exactly like the actual world except that all velocities are different. Likewise, if anti-haecceitism is true then there are no numerically distinct yet qualitatively identical possibilities. Since boosted worlds are qualitatively identical when spacetime has a Galilean structure, this once more means that there just are no boosted worlds—whether physically possible or not.

But this approach is mistaken: it is *physical* possibility that matters for underdetermination (and for indeterminism, but I will focus on underdetermination). The reason is that observation itself is a physical process. Therefore, whether a difference counts as empirical depends on what is physically possible. I will give two arguments for this claim. The first is that the very concept of detectability has a modal element. In particular, symmetry-variant quantities are undetectable *only* if the transformations under which they vary preserve physical possibility. The second argument is that underdetermination results from a kind of epistemic incoherence: the theory's evidence seems to undermine itself. This tension arises only when symmetry transformations preserve physical possibility. Again, this means that the existence of physically impossible worlds is of no concern.

I should note that although this second version of the objection is consistent with the quotations by Dasgupta and Pooley from the introduction, the first version is more natural. Indeed, I am not aware of anyone who has explicitly advanced the objection addressed in this section—apart, perhaps, from Dasgupta (2021). Nevertheless, it is useful to pre-empt the point and see where it fails.

3.1 Detectability

The first argument concerns detectability. The exact definition of detectability remains the topic of significant debate, but all hands agree that the following is a necessary condition on the successful measurement of a quantity Q:

> **Sensitivity**. If the value of Q were different, then a successful measurement of Q would have a different outcome.

For example, suppose that even if all velocities *v* were different from their actual values, any purported measurement of velocity would have the same outcome no matter what. It is uncontroversial that in that case velocities are undetectable.

The truth of the counterfactual in Sensitivity depends on which states of affairs are physically possible. Recall that on Lewis' (1979) theory of counterfactuals, it is of the utmost importance to avoid gross violations of the laws.[14] On the one hand, suppose that boosts are *not* physically possible, for instance because the laws say that the centre of mass of the universe is at absolute rest. What, then, is the closest possibility in which I have a different velocity? Presumably, it is not a boosted possibility—for such a possibility violates the laws. It is the possibility in which I have a different velocity with respect to, say, the surface of the earth. In that possible world, any purported measurement of my velocity *would* have a different outcome, since the measurement device itself will keep the same velocity. Therefore, velocities *are* detectable if boosts are physically impossible.[15] To put the point differently, in such a scenario absolute velocities have a dynamical role to play, such that it is possible to express one's velocity with respect to this dynamical role rather than merely relatively with respect to another body's velocity.

On the other hand, if boosts are physically possible then absolute velocities are not detectable. In particular, it seems plausible that in this case the closest possibility to ours in which my velocity is different from actuality is a boosted possibility: not only does a boosted velocity respect the laws, when spacetime is Galilean it also matches the actual world on all qualitative facts. It is therefore much closer to the actual world than the possibility in which I have a different relative velocity with respect to the surface of the earth. Because a boosted possibility is empirically equivalent to the actual world, any purported velocity measurement would have the same outcome. Sensitivity is violated; velocities are undetectable. Therefore, whether absolute velocity is undetectable depends on whether boosts are physically possible. Since it is the undetectability of symmetry-variant quantities that drives the problem of underdetermination, this means that that problem only arises when symmetries relate physically possible worlds.

3.2 Epistemic Incoherence

The second argument concerns the particular *kind* of undetectability that occurs in the presence of symmetries. There are many ways in which something can be undetectable. Some things are undetectable merely because we lack the requisite technology. Other things are undetectable because they are forever inaccessible to us, such as the distant past or beyond the event horizon. But even these are detectable in the weak sense that if one were to be able to travel to the end of time or the outer reaches of space, one could observe what was going on there.

Symmetry-induced undetectability is different. Here, it is the theory *itself* that tells us that symmetry-variant quantities are in principle undetectable. It is *because* we believe that classical mechanics is correct about absolute velocities that we must believe that such velocities are undetectable. Healey (2007) puts it well:

> If we believe Newton's theory as he interpreted it, then we believe there is a structure—the state of absolute rest—about which there is no way of obtaining reliable information by observation. This is not because of the contingent limitations of human sense organs, but because the theory countenances no physical process that discriminates that state from a host of others. [...] We can certainly entertain the belief that Newton's theory is

[14] I don't intend this argument to hang on Lewis' account of counterfactuals specifically; rather, it offers a simple illustration of how physical possibility matters to detectability.

[15] For a similar claim, see Jacobs (2021, §6.2).

> true as he interpreted it, and so also the belief that there is a unique state of absolute rest. But no matter how much evidence we obtained for Newton's theory, we would have no reason to hold this last belief. Consequently, we would have no reason to believe Newton's theory as he interpreted it. (Healey 2007, 117)

Put differently, theories in which symmetry-related models are taken to represent distinct physically possible worlds are in epistemic tension with themselves: any piece of evidence for the truth of such theories is *ipso facto* evidence that one can never discover their full truth.[16]

This epistemic incoherence arises only, however, when pairs of symmetry-related worlds are both physically possible. Compare the situation for classical mechanics to that for a theory in which boosted worlds are merely metaphysically possible. Suppose, for example, that there is a universal force that acts on bodies in proportion to their velocities. In that case, absolute velocities *are* detectable: they have an observable effect on the world. Nevertheless, for any world W in which this theory is true, there is another world W′ exactly like W except that all velocities are boosted. Because boosts preserve distances these worlds are empirically equivalent, yet W′ is not physically possible by the light of this theory's dynamics. If it is metaphysical rather than physical possibility that should concern us, this scenario would lead to the same problem of underdetermination as before. But that is clearly not the case. The non-Galilean invariant theory does provide us with evidence that absolute velocities exist, and even with the means to measure them. Therefore, evidence in favour of this theory does not undermine the theory's truth; it is not evidence that one cannot measure absolute velocities. The mere metaphysical possibility of boosted worlds does not cause any epistemic tension. Insofar as symmetries are worrisome because they lead to such tension, then, physical possibility is what counts.

One further lesson to draw from these considerations is that theses such as relationism or anti-haecceitism are in a sense 'overpowered'. In order to avoid symmetry-induced underdetermination, it suffices to rule out certain transformations as *physically* impossible—not impossible *per se*. This is not to say that there are no other reasons to prefer such metaphysical pictures, for example for reasons of parsimony. I will say more about this in the close of the paper.

4 Close

I have argued that Leibniz Equivalence places a restriction on the space of physically possible worlds, and that physical possibility is the relevant notion of possibility in discussions of detectability. It follows that endorsement of Leibniz Equivalence constitutes a solution to symmetry-induced underdetermination, contrary to recent objectors.

One may wonder whether it is always possible to simply *stipulate* that symmetry-related models represent the same possibility. Shouldn't such a stipulation follow from, or at least stay consistent with, the metaphysical interpretation of such models? It may now seem as if underdetermination is far too easily averted: just declare that pairs of models that threaten the determination of the theory by the empirical data are really representatives of the same class of internally equivalent possibilities under any representational convention. This will always suffice to avert a theory's underdetermination, since the theory's underdetermined features are by definition demoted to a merely external role. Perhaps this is a final sense in which Leibniz Equivalence does not suffice to solve the underdetermination problem.

[16] It seems that the incoherence discussed in Huggett and Wuthrich (2013) is of a similar kind.

I admit that one cannot always simply *declare* that certain pairs of models represent the same physical possibility (or, equivalently, declare that certain quantities are merely external). For a trivial example, consider a representational convention on which *all* of a theory's models represent the very same possibility. In that case, none of the theory's quantities are internal except perhaps the constants of nature. But that clearly conflicts with the fact that a theory, if it is *about* anything at all, is about the quantities that feature in its dynamics: quantities such as force, mass and acceleration that take on different values in different physically possible worlds. So, our representational conventions are constrained by the demand to make sense of our physical theories; and conversely the ways one can make sense of our physical theories are constrained by the representational conventions one has adopted.

If we now return to *symmetry*-related pairs of models, the attraction of Leibniz Equivalence lies exactly in the fact that theories are *not* about symmetry-variant quantities. This is the point of so-called 'symmetry-to-reality inferences' (Dasgupta 2016). In more detail, Wallace (2019b) shows that the dynamics of symmetry-variant quantities are effectively decoupled from the dynamics of the symmetry-*in*variant quantities; there is a sense in which the former have no effect on the latter. In the words of Baker (2022), such quantities are 'epiphenomenal'. It is this feature that allows one to consider them as external to the theory's subject domain.

Of course, one may still debate the exact conditions under which one is allowed to adopt Leibniz Equivalence. On one end of the spectrum, one could claim that it is possible to 'extract' a physical interpretation of the theory's models from a restriction on their representational capacities. This is the approach taken by Dewar's (2019) 'sophistication', which is an instance of the broader approach known as *interpretationalism*. On this view, one can always stipulate equivalences between models in order to define their common structure. On the other end, Møller-Nielsen's (2017) *motivationalism* holds that one can only identify symmetry-related models once one has a 'perspicuous metaphysical characterisation' of the underlying physical content in virtue of which they are physically equivalent.[17] On this approach, Leibniz Equivalence is merely a *desideratum* that motivates the search for a perspicuous metaphysical picture.

I therefore very much concur with the following remark of Teitel's:

> Suppose [Leibniz Equivalence] were true: we now know that we use each member of any equivalence class of [symmetry-]related solutions to model the same nomic possibilities. Yet that is all we have learned, and so this doctrine just raises a host of further questions: which nomic possibilities do they represent? And what are those possibilities like? […] At best, determining which differences are [symmetry-variant] is a preliminary step towards developing an adequate answer to the non-mathematical questions we're generally interested in. (Teitel 2021a, 10)

This is undoubtedly correct: Leibniz Equivalence *is* a merely preliminary step in the interpretation of a theory with symmetries. This does not mean that Leibniz Equivalence is consistent with underdetermination: we have seen that it is not. Rather, it means that there are non-trivial conditions on the warrantability of Leibniz Equivalence. This doesn't show that representational conventions are irrelevant to the question of underdetermination, as Teitel seems to believe. Quite the opposite: it means that representational conventions non-trivially interact with the exact metaphysical questions that Teitel is interested in.

[17] For Møller-Nielsen, a perspicuous picture is always available the symmetry-related models of a theory are isomorphic; if not, a reformulation of the theory is necessary. For further discussion of interpretationalism vs motivationalism, see Martens and Read (2020) and Jacobs (2022).

Finally, let me comment on an issue raised at the end of the previous section. The typical conclusion of a symmetry-to-reality inference is that symmetry-variant features are not real, yet to avoid underdetermination it suffices to consider symmetry-variant quantities as merely external. However, there are other reasons to support the symmetry-to-reality inference. One is simply the virtue of parsimony: if absolute position or intrinsic identity play no dynamical role, then surely a theory is better off without them. Another reason consists of a particular claim about fundamentality: that the internal quantities are more fundamental than the external ones. It is easy to denounce absolute velocities as external, but on an intuitive picture of motion absolute acceleration depends on absolute velocity. It took the development of Galilean spacetime to discover how acceleration could itself be fundamental. This explains why discussion has focused on 'overpowered' theses such as relationism or anti-haecceitism. Although these theses are stronger than necessary to have Leibniz Equivalence, they are required to provide a sensible metaphysical picture of the world. These constraints further complicate the task of the interpreter. Again, however, this does not mean that underdetermination is independent from Leibniz Equivalence, but only that the latter raises as many questions as it answers.

Acknowledgements: I would like to thank Adam Caulton, Henrique Gomes, Oliver Pooley, David Wallace and the audience at the University of Toronto's Logic and Philosophy of Science Group for helpful feedback and discussion.

Bibliography

Baker, D. J. (2010). Symmetry and the Metaphysics of Physics. Philosophy Compass, 5(12), 1157–1166. https://doi.org/10.1111/j.1747-9991.2010.00361.x

Baker, D. J. (2022). The Epiphenomena Argument for Symmetry-to-Reality Inference. http://philsci-archive.pitt.edu/20751/

Belot, G. (1995). New Work for Counterpart Theorists: Determinism. The British Journal for the Philosophy of Science, 46(2), 185–195. https://doi.org/10.1093/bjps/46.2.185

Belot, G. (1998). Understanding Electromagnetism. The British Journal for the Philosophy of Science, 49(4), 531–555.

Bradley, C., & Weatherall, J. O. (2020). On Representational Redundancy, Surplus Structure, and the Hole Argument. ArXiv:1904.04439 [Physics]. http://arxiv.org/abs/1904.04439

Brighouse, C. (1997). Determinism and Modality. The British Journal for the Philosophy of Science, 48(4), 465–481.

Brighouse, C. (2020). Confessions of a (Cheap) Sophisticated Substantivalist. Foundations of Physics, 50(4), 348–359. https://doi.org/10.1007/s10701-018-0228-2

Butterfield, J. (1989). The Hole Truth. The British Journal for the Philosophy of Science, 40(1), 1–28.

Dasgupta, S. (2011). The Bare Necessities. Philosophical Perspectives, 25(1), 115–160. https://doi.org/10.1111/j.1520-8583.2011.00210.x

Dasgupta, S. (2016). Symmetry as an Epistemic Notion (Twice Over). The British Journal for the Philosophy of Science, 67(3), 837–878. https://doi.org/10.1093/bjps/axu049

Dasgupta, S. (2021). Symmetry and Superfluous Structure: A Metaphysical Overview. In E. Knox & A. Wilson (Eds.), The Routledge Companion to Philosophy of Physics. Routledge.

Dewar, N. (2019). Sophistication about Symmetries. The British Journal for the Philosophy of Science, 70(2), 485–521. https://doi.org/10.1093/bjps/axx021

Earman, J. (2004). Laws, Symmetry, and Symmetry Breaking: Invariance, Conservation Principles, and Objectivity. Philosophy of Science, 71(5), 1227–1241. https://doi.org/10.1086/428016

Earman, J., & Norton, J. (1987). What Price Spacetime Substantivalism? The Hole Story. The British Journal for the Philosophy of Science, 38(4), 515–525. https://doi.org/10.1093/bjps/38.4.515
Fletcher, S. C. (2020). On Representational Capacities, with an Application to General Relativity. Foundations of Physics, 50(4), 228–249. https://doi.org/10.1007/s10701-018-0208-6
Greaves, H., & Wallace, D. (2014). Empirical Consequences of Symmetries. The British Journal for the Philosophy of Science, 65(1), 59–89. https://doi.org/10.1093/bjps/axt005
Healey, R. (2007). Gauging What's Real: The Conceptual Foundations of Contemporary Gauge Theories. Oxford University Press.
Huggett, N., & Wüthrich, C. (2013). Emergent spacetime and empirical (in)coherence. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 44(3), 276–285. https://doi.org/10.1016/j.shpsb.2012.11.003
Jacobs, C. (2020). Absolute Velocities Are Unmeasurable: Response to Middleton and Murgueitio Ramirez. Australasian Journal of Philosophy.
Jacobs, C. (2021). Invariance or Equivalence: A Tale of Two Principles. Synthese.
Jacobs, C. (2022). Invariance, intrinsicality and perspicuity. Synthese, 200(2), 135. https://doi.org/10.1007/s11229-022-03682-2
Lewis, D. (1979). Counterfactual Dependence and Time's Arrow. Noûs, 13(4), 455–476. https://doi.org/10.2307/2215339
Luc, J. (2023). The Unmeasurability of Absolute Velocities from the Point of View of Epistemological Internalism. Erkenntnis. https://doi.org/10.1007/s10670-023-00679-2
Martens, N. C. M., & Read, J. (2020). Sophistry about symmetries? Synthese. https://doi.org/10.1007/s11229-020-02658-4
Maudlin, T. (1988). The Essence of Space-Time. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1988, 82–91.
Melia, J. (1999). Holes, Haecceitism and Two Conceptions of Determinism. British Journal for the Philosophy of Science, 50(4), 639–664. https://doi.org/10.1093/bjps/50.4.639
Middleton, B., & Murgueitio Ramírez, S. M. (2020). Measuring Absolute Velocity. Australasian Journal of Philosophy, 0(0), 1–11. https://doi.org/10.1080/00048402.2020.1803938
Møller-Nielsen, T. (2017). Invariance, Interpretation, and Motivation. Philosophy of Science, 84(5), 1253–1264.
Norton, J. (2020). The Hole Argument Against Everything. Foundations of Physics, 50(4), 360–378. https://doi.org/10.1007/s10701-019-00258-y
Pooley, O. (2021). The Hole Argument. In E. Knox & A. Wilson (Eds.), The Routledge Companion to the Philosophy of Physics. Routledge.
Pooley, O., & Read, J. A. M. (2021). On the mathematics and metaphysics of the hole argument. The British Journal for the Philosophy of Science. https://doi.org/10.1086/718274
Roberts, B. W. (2020). Regarding 'Leibniz Equivalence'. Foundations of Physics, 50(4), 250–269. https://doi.org/10.1007/s10701-020-00325-9
Roberts, J. T. (2008). A Puzzle about Laws, Symmetries and Measurability. The British Journal for the Philosophy of Science, 59(2), 143–168. https://doi.org/10.1093/bjps/axn009
Ruetsche, L. (2011). Interpreting Quantum Theories. Oxford University Press.
Saunders, S. (2003). Physics and Leibniz's Principles. In K. Brading & E. Castellani (Eds.), Symmetries in Physics: Philosophical Reflections (pp. 289–307). Cambridge University Press.
Teitel, T. (2021a). How to Be a Spacetime Substantivalist. Journal of Philosophy.
Teitel, T. (2021b). What theoretical equivalence could not be. Philosophical Studies, 178(12), 4119–4149. https://doi.org/10.1007/s11098-021-01639-8
Wallace, D. (2019a). Isolated Systems and their Symmetries, Part I: General Framework and Particle-Mechanics Examples [Preprint]. http://philsci-archive.pitt.edu/16623/
Wallace, D. (2019b). Observability, redundancy and modality for dynamical symmetry transformations [Preprint]. http://philsci-archive.pitt.edu/16622/

Weatherall, J. O. (2018). Regarding the 'Hole Argument'. The British Journal for the Philosophy of Science, 69(2), 329–350. https://doi.org/10.1093/bjps/axw012
Williams, P. (2019). Scientific Realism Made Effective. The British Journal for the Philosophy of Science, 70(1), 209–237. https://doi.org/10.1093/bjps/axx043