Redundant Epistemic Symmetries

James Read[*] and Thomas Møller-Nielsen[†]

Faculty of Philosophy, University of Oxford,
Radcliffe Humanities, Woodstock Road, Oxford OX2 6GG, UK

Abstract

We undertake a detailed analysis of three 'epistemic' approaches to symmetries, due, respectively, to Ismael and van Fraassen, Caulton, and Dasgupta. Finding faults with each, we proceed to develop our own epistemic approach to symmetries. Having done so, we present a concern regarding all epistemic accounts: they render the notion of a symmetry transformation redundant as a tool for metaphysical theorising about scientific theories.

Contents

[*]james.read@philosophy.ox.ac.uk
[†]thomasmollernielsen@hotmail.com

1 Introduction

In recent years, there has arisen an explosion of literature on the nature and interpretation of *symmetry transformations* in scientific theories. On the one hand, authors have sought to provide a precise definition of such transformations; in light of the apparent failure of what Dasgupta [10, §5] calls *formal* and *ontic* definitions (a large number of the latter being presented in [2]), several (e.g. [8, 10, 17]) have moved to an alternative, *epistemic* conception of such transformations. (Very roughly, epistemic definitions of symmetry transformations involve conditions regarding the *observational indistinguishability* of symmetry-related solutions; not so for formal and ontic definitions.) On the other hand, there exists a debate regarding the *interpretation* of symmetry transformations: should symmetry-related models *invariably* be regarded as being physically equivalent, i.e. as representing the same possible world? [12, 20]

In this paper, our principal concern is with the definition of symmetry transformations. In §§3-5, we assess, respectively, the proposals for epistemic approaches to symmetry transformations due to Ismael and van Fraassen [17], Caulton [8], and Dasgupta [10], identifying problems with each. We then proceed in §6 to advance our own proposal for an epistemic definition of a symmetry transformation. This done, in §7, we arrive at the central concern of this paper: having constructed a general epistemic definition of a symmetry transformation, the notion of a symmetry becomes, in an important sense, *redundant* as a tool for metaphysical theorising about scientific theories. The reason for this is that, on such an epistemic approach, any claim that symmetry-related models should be regarded as being physically equivalent reduces simply to a mandate not to introduce undetectable structure into one's ontology, in line with Ockham's razor. In other words, without an *antecedent* notion of a symmetry transformation such that symmetry-related models are then *discovered* to be empirically equivalent, the notion of a symmetry has no novel interpretative role to play. Of course, whether one regards this as being a problematic feature of epistemic definitions of symmetry transformations is open for question—but what we wish to register in this paper is that the analogous issue does *not* arise for formal and ontic definitions of symmetries. While those approaches face problems of their own—in particular, what Dasgupta calls the "problem of inferential circularity" [10, p. 864]—they do not face this charge of conceptual redundancy. (Moreover, as we will discuss in §§6 and 7, it is not obvious to us that these alleged problems for non-epistemic definitions of symmetries cannot be avoided.)

Before we begin, it is worth flagging that in this paper we are not neutral on the question of the interpretation of symmetry transformations: we endorse a so-called *motivational* approach to symmetry transformations (cf. [20, 26]), according to which models of a given theory related by a symmetry transformation may not invariably be regarded *ab initio* as being physically equivalent; rather, such transformations at most *motivate us to seek* a clear ('metaphysically perspicuous') conception of physical reality according to which such models may indeed be so regarded. This is situated against the so-called *interpretational* approach, according to which models related by a symmetry transformation may invariably be regarded as physically equivalent (advocates of this interpretational approach include Dewar [12] and Saunders [28]). This distinction will be discussed further in due course.

2 Symmetries

In this section, we introduce three preliminary topics regarding symmetry transformations. In §2.1, we discuss the *model-theoretic*, or *semantic*, conception of scientific theories. In §2.2, we follow Dasgupta [10, §§5-6] in distinguishing *formal*, *ontic*, and *epistemic* definitions of symmetry transformations. In §2.3, we consider the *invari-*

ance principle, which states that only structure which is invariant between symmetry-related models of a given theory should be interpreted as having ontological import.

2.1 The Semantic Approach

On the *semantic conception* of scientific theories (see e.g. [32, 34, 35]), a theory is associated with a class of *models*. For a given theory $\mathcal{T}$, we take the most general class of associated models to be that of *kinematically possible models* (KPMs) $\mathcal{K}$, which consists in tuples of specified geometrical objects. For example, the KPMs of general relativity (GR) are picked out by all triples of the form $\langle M, g_{ab}, \Phi \rangle$, where M is a four-dimensional differentiable manifold, g_{ab} is a Lorentzian metric field on M, and Φ is a placeholder for the matter fields of the theory.

Classically, a theory $\mathcal{T}$, with KPMs $\langle M, O_1, \ldots O_n \rangle$ (where the O_i are geometrical objects), comes with a set of *dynamical equations* for the O_i. The KPMs of $\mathcal{T}$ in which the O_i obey those dynamical equations form a subspace $\mathcal{D} \subset \mathcal{K}$, the *dynamically possible models* (DPMs) of $\mathcal{T}$. For example, in the case of GR, only those triples $\langle M, g_{ab}, \Phi \rangle$ the geometrical objects of which satisfy the *Einstein equation*

$$G_{ab} = 8\pi T_{ab} \tag{2.1}$$

—the dynamical equation of the theory, which relates g_{ab} to the stress-energy tensor T_{ab} of the Φ—in *addition* to the dynamical equations of the Φ, are DPMs. Quantum mechanically, the story changes: some of the O_i in the KPMs of $\mathcal{T}$ are understood to be *operator-valued*; DPMs are picked out as those KPMs the geometrical objects of which satisfy certain *correlation functions* (for details, see [25, ch. 5]).

Models of a theory $\mathcal{T}$ may be *interpreted* as representing possible worlds. Sometimes, however, we may wish to interpret two or more distinct models as representing the *same* world. In that case, the space of KPMs $\mathcal{K}$ of $\mathcal{T}$ is partitioned into classes of *gauge-equivalent* models—which are interpreted as representing the same world—and the multiplicity of models representing the same world is an example of a *gauge redundancy*. In the case in which the interpretation of $\mathcal{T}$ leads to gauge redundancy, we *may* (but are not mandated to—see §4) construct a *reduced* space of models $\tilde{\mathcal{K}}$, in which gauge-related models are mathematically *identified*. This in turn induces a reduced space of DPMs, $\tilde{\mathcal{D}} \subset \tilde{\mathcal{K}}$.

2.2 Ontic versus Epistemic Definitions

The above is purely formal; there remains an outstanding question concerning *when* two models of a given theory $\mathcal{T}$ should be interpreted as representing the same possible world—i.e. be interpreted as manifesting a gauge redundancy, in the sense of §2.1. One ongoing debate—the subject of §2.3—regards the claim that two models of any given $\mathcal{T}$ may be regarded as physically equivalent (i.e. as representing the same possible world) when they are related by a *symmetry transformation*. In order for such a claim—and the surrounding debate—to have content, a definition of 'symmetry' in the context of scientific theories must first be provided. It is the purpose of this subsection to address some preliminary issues on this matter.

When considering the most appropriate definition of a symmetry transformation, it is valuable to distinguish—following Dasgupta [10, §§5-6]—*formal*, *ontic*, and *epistemic* definitions. Formal definitions "define [the notion of symmetry] in purely formal, set-theoretic terms" [10, p. 861]. Dasgupta goes on to criticise formal approaches, for two reasons. The first is that no candidate formal definition appears to be extensionally adequate: some such candidate definitions would appear to map observationally distinguishable (and hence patently not physically equivalent) solutions to one another; other such candidate definitions would seem to rule out what appear to be *bona fide* symmetry transformations as constituting 'genuine' symmetry transformations (e.g. absolute velocity boosts in Newtonian gravitation theory). The second, more general problem that Dasgupta raises for formal definitions is that any such definition "must imply that given any set of laws, any two systems related by a symmetry of those laws will be observationally equivalent. And it is (to put it mildly) extremely hard to see how any purely formal definition could have this consequence" [10, p. 861].

We do not wish to dwell on these criticisms of formal definitions of symmetry transformations—though, anticipating some of our results to be presented later in this paper, we should say that it is possible that the approach to empirical equivalence delineated in §6 may be applied in order to overcome Dasgupta's second above concern; moreover, it is questionable whether Dasgupa's concern that some candidate formal definitions of symmetries map observationally distinguishable solutions to one another truly need be considered problematic, in light of §7. Given this, turn now to ontic definitions of symmetry transformations. Taking as given that "The general idea is to define a symmetry of a law to be a function that preserves the law and also preserves ... features F" [10, p. 862], according to ontic definitions of symmetries, we are to "restrict our attention to physical features like distance, mass, charge, spin, and so on. ... The result is a definition of 'symmetry' that requires a symmetry to preserve the laws and preserve certain privileged physical features" [10, p. 862].[1]

One particularly popular ontic definition of a symmetry transformation (found in e.g. [13, 27]) is the following: A symmetry transformation is a map from KPMs of a theory $\mathcal{T}$ to KPMs, such that DPMs are always mapped to DPMs.[2] As Belot makes clear, there are profound difficulties with this definition. To bring this out, he lays out the following two principles: [2, p. 3]

D1: The symmetries of a theory are those transformations that map solutions of the theory to solutions.

D2: Two solutions of a theory are related by a symmetry transformation only if they are physically equivalent.

(**D2** will be discussed in depth in §2.3.) Belot argues—in our view correctly—that **D1** is too broad as a definition of a symmetry transformation: "Ordinarily, symmetries of theories are hard to come by. But some remarkable theories have atypically

[1]We are somewhat sceptical that the distinction between formal and ontic approaches to symmetries is as precise as Dasgupta suggests. For instance, are Lagrangian symmetries purely formal—given that they can be given an entirely precise, mathematical definition—or are they ontic—given that such symmetries will also tend to leave invariant certain privileged features of a given physical system?

[2]Dasgupta [10, p. 863] takes this definition to count as ontic rather than formal, for in this case the physical laws are preserved. However, since a symmetry transformation typically is supposed to be cashed out in terms of *some* map effected on the space of KPMs of the theory in question, such that DPMs are taken to DPMs, we question whether 'preserving the laws' is sufficient for a definition of a symmetry transformation to qualify as ontic, rather than formal. Nevertheless, nothing will hinge upon this issue, and in the remainder of this paper such concerns are set aside.

large symmetry groups. The definition above effaces this sort of distinction between theories. For if we allow arbitrary permutations of the solutions of a theory to count as symmetries, then the size of a theory's group of symmetries depends only on the size of its space of solutions" [2, p. 6]. In light of this problem, Belot then proceeds to consider more nuanced definitions of symmetry transformations [2, §§3-4] (for example, 'classical symmetries' [2, p. 8], 'generalised (Noether) symmetries' [2, p. 10], 'nonlocal symmetries' [2, p. 11], 'variational symmetries' [2, p. 12], and 'Hamiltonian symmetries' [2, p. 13]), but argues that all such ontic definitions come into conflict with **D2**. Ultimately, Belot takes this as reason to reject **D2**, rather than these definitions of symmetry transformations; as will become apparent over the course of this paper, we concur with this verdict.

For Dasgupta [10, §5.2], there are certain important reasons to be wary of ontic definitions of symmetries:

> The objection is that they get the order of justification backwards: we often use premises about symmetries in order to work out which physical features fix the data, so we cannot at the same time define symmetries to be those operations that preserve features that fix the data. [10, p. 865]
>
> The implicit ingredient ... is that symmetries must preserve the appearances—in some sense to be made clear. If we build this into our definition of symmetry, it will follow by definition that given any set of laws, any two situations related by a symmetry of those laws are observationally equivalent. This (remember) is what we need in order to underwrite the argument that variant features are undetectable. This is the epistemic approach to defining 'symmetry'. [10, p. 866]

Given this problem of "inferential circularity" with ontic definitions of symmetries—a problem which we will consider more explicitly in §§6 and 7—Dasgupta concludes that symmetry must, in fact, be an *epistemic* notion.[3] According to epistemic defini-

[3] In fact, Dasgupta argues that symmetry is an epistemic notion 'twice over'. The first sense in which this is so is that above: for Dasgupta, the definition of a symmetry transformation proceeds (in part) in terms of epistemic criteria. The second sense in which this is so is that, for Dasgupta, symmetry transformations are related to the Ockhamist norm that undetectable structure should be excised from our scientific theories.

tions of symmetries, "we allow [features preserved between symmetry-related models] F to include epistemic or observational features such as 'looking red to me' or 'appearing from my perspective to be two feet away' " [10, p. 862]. We appraise the problem of inferential circularity in §§6 and 7; in the meantime, we discuss in §§3-5 some specific proposals for epistemic definitions of symmetry transformations.

2.3 The Invariance Principle

We turn now to matters regarding the interpretation of symmetry-related models of a given theory. The most important thesis in this regard, from the point of view of contemporary literature on this topic, is **D2**: that two solutions of a theory related by a symmetry transformation may legitimately be regarded as physically equivalent. Such a claim embodies what we follow [20,26] in calling the *interpretational* approach to symmetry transformations (cf. §4); contemporary advocates of (some suitably close version of) **D2** include Dewar [12] and Saunders [28,29]. Sometimes, **D2** is packaged alongside what Saunders [29, p. 453] calls the *Invariance Principle* (**IP**):

IP: Only quantities invariant under all symmetries of the theory in question should be regarded as being 'real'—i.e. be interpreted as having ontological import.

Note, however, that **D2** and **IP** are logically *independent*: two models of a given theory being related by a symmetry transformation—and therefore, by **D2**, being regarded as physically equivalent—does not imply that the interpretation of those models should proceed in terms of their invariant structure; on the other hand, two symmetry-related models being afforded an interpretation proceeding exclusively in terms of their invariant structure does not imply that those symmetry-related models must be regarded as being physically equivalent, for the invariant quantities between those models may be afforded a *different* interpretation in each case.

Having introduced the model-theoretic framework used in this paper; distinguished between formal, ontic, and epistemic definitions of symmetry transformations; and presented **D2** and **IP**, we turn now to the first of three prominent epistemic approach to symmetry transformations found in the literature—this due to Ismael and van Fraassen [17]. After evaluating this approach, we then consider other epistemic approaches to symmetries (§§3-5); advocate the motivational approach to symmetries (introduced above) (§4); present our own epistemic approach to symmetries (§6); and raise some general concerns regarding epistemic approaches to symmetries (§7).

3 Approach I: Ismael and van Fraassen

Ismael and van Fraassen's central concern is to provide an account of how and why drawing symmetry-based ontological inferences is justifiable. They argue that the 'qualitative indiscernibility' of a theory's solutions should play a crucial role in any such account, construing 'qualitative properties' as those that are "directly observationally accessible to the observer," and are "distinguishable by ... a gross discrimination of colour, texture, smell, and so on" [17, p. 376]. It is worth noting that Ismael and van Fraassen are careful to contrast qualitative properties with those that are merely 'measurable' in their sense: such measurable properties are those that are able to "make some discernible impact on gross discrimination of colour, texture, smell and so on ... [no] matter how attenuated the connection is, how esoteric the impact, or how special the conditions under which it can be discerned" [17, p. 376].

Ismael and van Fraassen summarise their proposal regarding symmetry transformations as follows: [17, p. 380]

> *...[W]e submit that it is precisely the qualitative-structure-preserving symmetries of the laws that are indicative of the presence of superfluous theoretical structure and should always be interpreted as trivial*. (Emphasis in original)

The thought would seem to be this. Take the space of DPMs $\mathcal{D}$ of a given theory $\mathcal{T}$; determine which solutions are qualitatively indiscernible in the sense of their involving the exact same distribution of 'directly observable' quantities; take the relevant set of solution-preserving transformations on the space of KPMs $\mathcal{K}$ of $\mathcal{T}$ to be just those that map qualitatively indiscernible solutions to one another; and then take solutions related to one another in this way to represent the same physical state of affairs.[4]

In assessing this account, the first obvious question to ask is this: why are Ismael and van Fraassen so careful to stress the distinction between quantities or properties that are 'qualitative' or directly observable, and those that are merely 'measurable'? Moreover, why do they think that this is a distinction that is relevant to the **IP** (in the sense of being indicative of "the presence of superfluous theoretical structure")? Unfortunately, the answer to both of these questions is rather difficult to discern from what Ismael and van Fraassen write; indeed, they often write as if it *is* the measurable/unmeasurable distinction, rather than the qualitative/nonqualitative distinction, that is relevant here. (E.g., "...[O]ur main topic: superfluous structure will align with the presence of *unmeasurable* quantities in the theory's world picture. [...] To sum up: we are going to connect superfluous structure with the presence of *unmeasurable* quantities." ([17, pp. 376, 378], our emphasis).) They do on occasion (e.g. [17, pp. 376, 378]) suggest that the distinction is crucial because "what is measurable/unmeasurable cannot be read off directly from the theory," and that "[w]e need to make use of what is observable [i.e., qualitatively discernible] in order to make this distinction." But, even assuming the correctness of this claim, the question arises: Why should the fact that certain quantities can be 'read off' more-or-less 'directly', rather than 'indirectly', from a given theory's formalism, have any bearing on what kind of quantities superfluous structure-indicating transformations should necessarily preserve? To be sure, which quantities are directly observable might be *easier to determine* than which quantities are merely measurable. But why should this have any particular bearing on the **IP**, or the issue of which symmetry transformations on $\mathcal{K}$ among the $\mathcal{D}$-preserving ones are the *right* ones? Ismael and van Fraassen do not provide an answer to this question in their paper; nor is it easy to imagine in what a satisfactory answer could consist.

[4]Note that, unlike e.g. Dasgupta [10], Ismael and van Fraassen do not *define* a 'symmetry transformation' to be any empirical-data-preserving mapping. Rather, they highlight that the symmetry transformations which are *relevant* to symmetry-to-reality based inferences are those which point to the existence of undetectable structure—i.e., are those symmetries which relate empirically equivalent states of affairs [17, p. 379]. We fully agree with Ismael and van Fraassen on this point. Since these authors do not explicitly define symmetries in epistemic terms, the central concern of this paper—that in so doing the notion of a symmetry is rendered redundant as a tool for metaphysical theorising about scientific theories—does not apply. Our central concern regarding Ismael and van Fraassen's account will, rather, centre upon their distinction between the qualitative and the measurable—see below.

A second concern regarding this account is the following. Given Ismael and van Fraassen's explicit admission that "many measurable quantities will be non-qualitative" [17, p. 376], their account seems to imply that $\mathcal{D}$-preserving transformations which map *measurably distinct*, but nevertheless 'qualitatively indiscernible', solutions to one another could plausibly be interpreted as revealing the presence of superfluous structure. But such a consequence is patently absurd. Now, of course, the obvious way to remedy this defect in their account would be to connect the presence of superfluous structure to the existence of unmeasurable quantities—and, indeed, we noted above that this is how Ismael and van Fraassen often seem to frame their view—but then it would seem that the qualitative/nonqualitative distinction has lost its primary relevance for their account: it would be symmetries that preserve all *measurable* structure that would be indicative of surplus structure, with the qualitative/nonqualitative distinction having no substantive role to play (other than, perhaps, insofar as qualitative quantities help us to 'get a fix on' what is measurable according to a given theory).

Our third concern regarding Ismael and van Fraassen's account is perhaps the most interesting.[5] If we agree with Ismael and van Fraassen's proposal that one can 'directly observe' certain quantities, then it would appear to follow almost ineluctably that such quantities must remain 'directly observable' even after a transition to a successor theory. The history of physics, however, would seem to suggest a very different lesson. Thus, for instance, in the transition from Newtonian mechanics to relativity theory, it is plausible to think that the 'directly observable' distance between two physical events transitioned from their spatial distance *simpliciter*, to their 'spatial distance' relative to a particular frame of reference. The move to a new theory, and a new associated set of transformations, might well yield the conclusion that quantities that we previously thought we could straightforwardly and unproblematically detect are, in fact, not detectable or indeed not real after all. The notion that we 'directly observe' certain quantities *simpliciter*, then, *pace* Ismael and van Fraassen, is one that is extremely difficult, if not impossible, to square with the history of physics. We return to this matter in §6.

[5] This point has been noted explicitly by Dasgupta [10, p. 867], while Saunders [28, p. 300] also appears to make an essentially identical observation.

Despite these faults, in our view Ismael and van Fraassen latch onto something important in their epistemic approach to symmetry transformations, and their attempted explanation and justification of **IP**-based inferences. For we, too, think that issues to do with observation and, more specifically, detection and measurability are essential in justifying such inferences. These matters are brought out explicitly in the remaining sections of this paper.

4 Approach II: Caulton

The basic structure of Caulton's proposal regarding symmetry transformations is the following. One begins with the space of DPMs $\mathcal{D}$ of one's theory, with each element being endowed initially with what we might call a 'minimal' interpretation: this is Caulton's "first phase of the interpretative process" [8, p. 160]. This minimal interpretation fixes a determinate representation relation between each model and the empirical phenomena. Thus, given this representation relation, one is able to determine which models of one's theory are empirically distinguishable, or empirically indistinguishable, solutions, and by the same token one is able to determine which solutions are compatible with the empirical phenomena so far observed.

This minimal interpretation is, however, entirely non-committal on the issue of which models represent the same *physical* state of affairs. Fixing this relation of physical equivalence among models is the job of Caulton's "second phase of the interpretative process" [8, p. 160]. To this end, Caulton distinguishes what he calls "analytic symmetries" from "synthetic symmetries"[6] [8, §2.3]: the crucial point is that while synthetic symmetries relate physically *distinct* solutions of the theory, analytic symmetries instead relate solutions that represent the *same* physical state of affairs.[7] But how is one to determine which solutions are related by an analytic symmetry, without having antecedently determined which quantities of the theory are genuinely real?[8]

[6]The obvious analogy here, as Caulton notes, is between analytic and synthetic propositions.

[7]Caulton's definition of a "synthetic symmetry" is in fact more nuanced: he distinguishes between "synthetic symmetries of the first kind" (which fail to leave all genuine physical quantities invariant) and "synthetic symmetries of the second kind" (which preserve all genuine physical quantities only for some proper subset of a theory's solutions) [8, §2.3]. This subtlety need not concern us here.

[8]Essentially this same problem of "inferential circularity"—to use Dasgupta's useful label—in de-

The answer Caulton proposes is straightforward: one should regard as many transformations among the theory's solution space as analytic as is possible without compromising the theory's (assumed) empirical adequacy. As Caulton concisely puts it, one should "*maximise the analytic symmetries, subject to empirical adequacy*" [8, p. 160].

An initial worry one might have regarding Caulton's proposal—one which Caulton himself raises—is whether it is tantamount to "a form of verificationism" [8, p. 161]: more specifically, whether it is compatible with a form of realism that is not *forcibly* committed to regarding observationally equivalent models as representing the same physical state of affairs. Such a problem would seem to arise because, if it is indeed a constitutive feature of the suggested interpretative process that one is to "maximise" the number of analytic symmetries, subject to the (apparently, sole) condition of empirical adequacy, then it would appear to follow—by interpretative *fiat*—that one is unable to make sense of empirically indistinguishable but nevertheless physically distinct solutions: for the second phase of the interpretative process would appear to dictate that we *must* regard all empirically indistinguishable solutions as representing the same physical state of affairs.

In reflecting on this concern, it is useful to appeal to a distinction drawn in [20], between *interpretational* and *motivational* approaches to symmetries (cf. §2.3). Read as an interpretationalist, Caulton is claiming that empirically equivalent, symmetry-related solutions must *always* be taken to represent the same physical state of affairs. As noted, this seems to be a straightforward consequence of Caulton's "second phase of the interpretative process": that one should maximise the analytic symmetries of one's theory while respecting the theory's assumed empirical adequacy. According to a second construal of Caulton's proposal, however, one is merely inherently *motivated* to regard empirically equivalent solutions as representing the same physical state of affairs, but one is not *necessarily* committed to—or, indeed, warranted in—regarding such solutions as being physically equivalent.

On the first construal of Caulton's account, it appears inescapable that the proposed method will leave one unable to regard solutions that are empirically indis-

termining a given theory's relevant solution-preserving transformations has been independently noted by Nozick [21, p.79], Ismael and van Fraassen [17, p. 386], Debs and Redhead [11, p. 66], Swanson and Halvorson [33, p. 7], and, most perspicaciously, by Dasgupta himself [10, §5.3].

tinguishable as representing distinct physical possibilities. On the second construal, on the other hand, it seems easily escapable—but then the second phase of the interpretative process is *not* correctly described as a process in which one maximises the "analytic symmetries" of one's theory: rather, what one is doing during this process is working out which (empirically equivalent) solutions one may legitimately regard as representing the same physical situation. *Analytic* symmetry would therefore not be the apt term according to this second construal of Caulton's position: for on this construal interpretative practice would not in general require that the solutions in question be regarded as physically equivalent.

To illustrate the distinction between these two possible construals of Caulton's proposal, consider Newtonian gravitation theory (NGT) set in Newtonian spacetime. This theory has KPMs picked out by tuples $\langle M, t_{ab}, h^{ab}, \sigma^a, \rho, \phi \rangle$, where M is a 4-dimensional differentiable manifold; t_{ab} is a temporal 'metric' field of signature $(1, 0, 0, 0)$; h^{ab} is a spatial 'metric' field of signature $(0, 1, 1, 1)$; σ^a is a timelike vector field the integral curves of which represent the worldlines of the persisting points of absolute space; and ρ and ϕ represent the matter density and the gravitational potential field, respectively.[9] The symmetry group of NGT includes not only transformations corresponding to time-independent velocity 'boosts' of solutions' total matter content, but also transformations corresponding to time-dependent translational accelerations of such content (so long as the gravitational potential field is also appropriately transformed). Thus, read 'naïvely', the symmetries of this theory include transformations that map solutions to solutions that represent physically distinct, but nevertheless empirically indistinguishable, states of affairs in which a given material system is:

1. Force-free and stationary with respect to absolute space.
2. Force-free and moving at constant absolute velocity.
3. Absolutely accelerating under a gravitational force-field.

[9]Canonical presentations of Newtonian spacetime (e.g. [15, §III.1] and [13, §2.5]) take the affine connection as primitive. We find such presentations unsatisfactory for historical reasons: they threaten to make the move to Galilean spacetime seem almost trivial, and the associated timelike vector field superfluous. For more on this point, see [22, §§4.4-4.5].

According to the first construal of Caulton's proposal (which we take to be his actual position),[10] we may (and, it seems, should) take all of these mathematically distinct models as *in fact* representing the same physical state of affairs. On the second construal of Caulton's proposal, however, we are merely *motivated* to regard all such models as representing the same physical state of affairs—the motivation arising from the general Ockhamist principle that, other things being equal, our preferred scientific theories should not allow for physically distinct but nevertheless empirically equivalent solutions. According to this second construal of Caulton's position, absent a metaphysically perspicuous characterisation of the reality underlying the symmetry-related models in question (as is achieved by moving to e.g. Newton-Cartan theory[11]), we have no choice but to regard such solutions as representing physically distinct states of affairs.

Thus, there are two distinct ways of understanding Caulton's 'analytic symmetries'—as possessing (i) *interpretational*, or (ii) merely *motivational*, force. Setting Caulton's own views to one side, which of (i) or (ii) is the more plausible? We incline towards the latter. There is nothing, after all, obviously absurd about admitting in principle undetectable facts into one's ontology; nor is there any obvious reason why we should always be capable of discovering a theory (or a perspicuous characterisation of a theory) which explains such solutions' empirical equivalence in terms of their fundamental physical equivalence; nor is there even any obvious way of guaranteeing that there will always be such a theory waiting in logical space to be discovered.

There are also important explanatory matters to consider. For although the Newtonian who adopts the merely motivational construal of symmetries might indeed be committed to there being facts beyond one's epistemic grasp, she nevertheless has a perfectly good explanation as to *why* such facts are epistemically inaccessible: they are inaccessible precisely because the world is in fact truly and accurately represented by tuples of the form $\langle M, t_a, h^{ab}, \sigma^a, \rho, \phi \rangle$, and because all any Newtonian observer ultimately has direct empirical access to are the instantaneous relative distances and relative velocities between material entities.[12] For such a Newtonian, then, the em-

[10]Caulton has privately confirmed to us that he subscribes to this view.

[11]See [26, §3.2] for detailed discussion of this manoeuvre.

[12]As Stein notes [30, pp. 156-7], this is a simplification: instantaneous quantities are never *directly* empirically accessible. Rather, determining their values is an indirect process, and occurs over a finite

pirical phenomena underdetermine the genuine physical facts; however, the theory itself is able to provide a perfectly transparent explanation of the reality behind the phenomena in terms of which the underdetermination can be understood.

The Newtonian who adopts the interpretational construal of symmetries, however, would appear to lose this explanatory transparency: she knows *that* she must regard the symmetry-related solutions in question *as* 'physically equivalent', but the reality in terms of which this physical equivalence is to be understood may remain opaque to her; she is offered no explanation as to *how* such physical equivalence is to be construed, or how it could even be said to arise.

Our views in this vicinity also diverge from those of Caulton in another important respect. To illustrate this difference, it will help to quote Caulton's own summary of his account, which comes near the end of his paper: [8, p. 161]

> We proceed in two phases. During the first phase we set up representational links between the theory and the observable portion of the physical world, under the assumption that the the theory is empirically adequate (or similar). In the second phase, we maximise the theory's analytic symmetries, taking advantage of the representational links forged in the first phase so as not to compromise empirical adequacy. *The result is an interpretation for the theory that prompts appropriate reform towards a* new *formalism, in which the physical properties and relations, including the unobservable ones, are transparently represented without redundancy.* (Our emphasis.)

We disagree with the final, italicised claim in the above. More specifically, we disagree with the suggestion that *all* symmetry transformations should be construed as prompting a *reformulation* of the relevant physical theory.

It is important to distinguish this claim—that symmetry transformations invariably prompt a mathematical reformulation of the relevant theory—from the 'motivational' view of symmetries discussed (and endorsed) above. According to the motivational view of symmetries, one is at best only motivated to regard symmetry-related

period of time. In this regard, see also [5–7].

models as physically equivalent; moreover (and more specifically), one is justified in regarding such models as physically equivalent only insofar as one is in possession of a metaphysically perspicuous characterisation of the reality underlying such models. However—and this is the crucial point to note—we believe that one *can*, in fact, be in possession of such a metaphysically perspicuous characterisation of the reality underlying symmetry-related models *even in the absence* of a mathematical formulation of the theory which removes the relevant representational redundancy.

When is this the case? We think the answer is straightforward: when such models are *isomorphic*, or—equivalently—when they are naturally understood as representing at most haecceitistically distinct possible worlds. Thus, we claim, neither the time-independent translational ('Leibniz shift') symmetry of NGT, nor the diffeomorphism invariance of general relativity, *by themselves* motivate any mathematical reconstrual of the respective theories. This is because we believe there is a perfectly transparent, anti-haecceitist way of understanding such models' representational equivalence even in the absence of such a mathematical reformulation. This view denies that there are any primitive, singular (haecceitistic) facts about spacetime points which would even allow for a coherent distinction between shifted or diffeomorphically-related scenarios to be drawn; thus, shifted models in NGT, and diffeomorphically-related models in GR, are understood as representing the same physical state of affairs.[13]

In sum: *Pace* Caulton, it is only when symmetry-related models are *not* isomorphic that one is prompted to mathematically reformulate the theory in question, such that the two (distinct) original symmetry-related models of the original theory correspond (up to isomorphism) to the same model of the new theory, in terms of which the ontology of the original two models is to be understood.

[13] In the spacetime context, such a view is known as 'sophisticated substantivalism' (for presentations, see e.g. [3, 16, 18, 28]); a more general such anti-haecceitist view is the 'moderate structuralism' of Esfeld and Lam [14]. In recent work, Dasgupta [9] has claimed not to understand sophisticated substantivalism. We are sceptical as to whether there is anything that can be said that is dialectically effective against such confessions, other than to note that sophisticated substantivalism is *not* (to use Dasgupta's useful terminology, and as we construe it) a "bare modal" thesis: that is, it is not the "mere assertion" that worlds cannot differ purely over which spacetime points are playing which qualitative roles. Rather, such anti-haecceitism is supposed to be a *consequence* of moderate structuralism. Plausibly, Dasgupta would claim not to understand this modest structuralist view, to which all we can say in response is that the view seems perfectly clear *to us*.

5 Approach III: Dasgupta

Dasgupta, more flat-footedly and straightforwardly than Ismael/van Fraassen and Caulton, subscribes to the epistemic conception of symmetry transformations, according to which symmetries map solutions of a given theory to empirically indistinguishable solutions. Unlike Caulton, but like ourselves, he seems to subscribe to the motivational, rather than the interpretational, construal of symmetries: [10, pp. 853-4]

> ... [W]e can draw the conclusion of the inference [*viz.*, that a given theory's relevantly 'variant' quantities are not real] only when we have the alternative theory in hand and have shown that all else is equal. This explains why it was rational for Newton to believe in absolute velocity even though he knew that it was variant in [NGT] and undetectable. The reason this was rational for him was that he had no good alternative theory to hand. He had good reason (his bucket argument) to think that relationalism was not empirically adequate. And relationalism was the only alternative view he knew of (he was not aware of Galilean space-time structures in which there is a well-defined feature of absolute acceleration, as required by his bucket argument, but no absolute velocity). So for Newton, all else was not equal and he was rational to believe in absolute velocity.

We believe that this quote indicates Dasgupta's commitment to motivationalism because, as the quote makes clear, Dasgupta would appear to agree that symmetries at most only motivate (but do not invariably legitimate) one to regard symmetry-related solutions as physically equivalent: for Dasgupta, "all else must be equal" and we must have "a good alternative theory in hand" before we may legitimately regard (e.g.) boosted solutions in NGT as physically equivalent. That said, Dasgupta's view differs from our own—and concurs with Caulton's—insofar as he, too, believes that symmetries invariably motivate a mathematical reformulation of the theory in question, in order to eliminate its representational redundancy. As he goes on to write in a footnote: [10, fn. 23]

> There is a lesson here for contemporary structuralists, such as Ladyman

> and Ross (2007, 130), who take the fact that diffeomorphisms are symmetries of general relativity to suggest that "[t]here are no things. Structure is all there is". For it is not enough to note that individual points of the manifold are variant features and declare that they are therefore not real. That would be analogous to Newton declaring that there is no such thing as absolute velocity without a genuine alternative theory in hand, a move that we would rightly have regarded with suspicion. To motivate structuralism, one must present a clear theory of the fundamental structure of the material world without making reference to regions of the manifold, a theory that does well on other theoretical virtues such as simplicity, elegance, and so on. But contemporary structuralists tend not to present such a theory.

We concur with Dasgupta's suggestion that structuralists—*insofar as they wish to dispense with spacetime points* qua *primitive entities*—should be motivated to seek a mathematical alternative to GR which does not explicitly quantify over points of the spacetime manifold. We also concur with the suggestion that the diffeomorphism invariance of GR should motivate us to articulate a clear conception of the reality supposedly underlying diffeomorphically-related models. However, we do not concur with the suggestion that the mere existence of this symmetry should motivate us to develop an 'alternative theory'—if by 'alternative theory' one means a *formally* distinct theory. The reason for our disagreement is (to repeat) because diffeomorphism-related models are *isomorphic*: they represent at most haecceitistically distinct possible worlds. And, as mentioned above, there is a perfectly transparent way of understanding such isomorphic models' physical equivalence which by no means requires any mathematical modification or reformulation of GR: namely, by adopting an anti-haecceitist ('sophisticated substantivalist') conception of spacetime points.

It should be noted that this fact by itself does not entail that Dasgupta is not correctly described as a motivationalist. Rather, we think it is better to describe him as a motivationalist who does not subscribe to our specific form of motivationalism, as he has a different conception of what would constitute a 'metaphysically perspicuous' characterisation of the reality (or a "clear theory") in terms of which the physical equivalence of isomorphic symmetry-related models can be understood: more specifically, Dasgupta would not regard sophisticated substantivalism as counting as a "clear theory" explaining how and why (e.g.) diffeomorphic models in GR can be identified.

Dasgupta also differs from Caulton and ourselves in another important respect: he thinks that more needs to be done to explicate properly the notion of 'empirical indistinguishability' at play in considerations regarding symmetry transformations. He reasons: [10, pp. 866-7]

> Rather than being a primitive relation, it is natural to think that the relation of observational equivalence holds between two structures in virtue of their intrinsic properties. So a better definition [of 'symmetry'] would identify the intrinsic properties that make for observational equivalence, and then define a symmetry to be a transformation that (in addition to preserving the laws) preserves them.

What could these "intrinsic properties that make for observational equivalence" be? Dasgupta argues that they must not be *physical* features, for to identify them as such would be to fall into the trap of "inferential circularity": that is, of merely stipulating, without any apparent antecedent justification, that some particular physical feature must be preserved under a theory's **IP**-relevant transformations and then concluding on that very basis that this (*ex hypothesi* invariant) feature is genuinely real (more on this problem in §§6 and 7). Rather, he argues, any satisfactory definition of the relevant notion of symmetry, and the associated notion of observational equivalence, must be spelled out in what he calls "epistemic terms": that is, in terms which ultimately

> do not depend on the underlying metaphysics, such that we can be in a position to know whether two structures are observationally equivalent prior to knowing anything (via symmetry-to-reality reasoning) about the metaphysics of our world. [10, p. 867]

Dasgupta then goes on to sketch two ways in which a proper 'analysis' of the notion of observational equivalence could go: one in terms of a (suitably rigorised) notion of 'how things look'; the other in terms of the notion of 'observation sentence' drawn from Quine [23, 24].[14] Moreover, although Dasgupta [10, p. 871] is careful to stress that he does not fully endorse either such 'analysis' of observational equivalence (they "are merely ... approximation[s] of what a completed epistemic definition of symmetry might look like"), he is nevertheless explicit that "the legitimacy of symmetry-to-reality reasoning depends on our making out some notion of observational equivalence along these lines."

It is this latter conclusion with which we disagree. In particular, we do not agree that the notion of 'observational equivalence' *must* necessarily be amenable to an analysis of the kind Dasgupta considers in order for **IP**-style 'symmetry-to-reality' reasoning to be justified. For to fail to offer an 'analysis' of observational equivalence in terms of (e.g.) Quinean observation sentences does not seem to be equivalent to taking the notion of observational equivalence as a 'primitive' or unanalysable relation. For instance, in the Newtonian context, we *know* what accounts for the apparent 'observational equivalence' of the relevant symmetry-related states of affairs: it is the fact that they all share the same pattern of instantaneous relative distances and velocities instantiated by their respective material systems. It is *these* 'intrinsic properties', coupled with the (as it turns out, false) assumption that all we can (more-or-less) 'directly observe' are such instantaneous material distances, from which the noted observational equivalence may be inferred. One need never resort to any notion of 'observation sentences', nor to any rigorised notion of 'how things look', in this inferential process: these notions play, at best, a superfluous role.[15]

But is this not equivalent to committing Dasgupta's noted fallacy of inferential circularity? We do not believe so (cf. §§6 and 7). The assumption that relative spatial distances are all that we can truly detect is, after all, defeasible: its implicit assumption

[14]The basic idea behind Quine's analysis is that two worlds are empirically equivalent just in case they assign ('peg') the same observation sentences—roughly speaking, those sentences to which a generic subject will be immediately disposed to assent to or dissent from when appropriately neurally stimulated—to each spacetime point throughout the worlds' respective histories. The proposal is variously problematic. (For instance: Does it make sense to assign observation sentences to *points* in spacetime, given that human beings are not zero-dimensional objects? Does it make sense to speak of 'pegging' observation sentences to 'place-times' in which human observers could not survive? And which particular subject's assent/dissent to neural stimuli should we take as canonical anyway? Etc.)

[15]Of course, it might be *nice* to have some such 'formal', or suitably rigorised, or 'mathematicised', definition of observation equivalence. Our point is simply that there is no *philosophical* pressure to provide such an account.

of a unique global foliation of the spacetime manifold is dispensed with in relativity theory. The fact that what we take ourselves to 'directly detect' is itself open to revision on the basis of theoretical innovation is just one aspect of the practice of science, and of what van Fraassen [34, §3.7] has dubbed the "hermeneutic circle" more generally.

In other words, theories *themselves* are ultimately the best guides we have as to what we take ourselves to observe, and what we take ourselves to observe in turn provides epistemic support (or refutation) of those same theories. Taking the relevant solution-preserving transformations to invariably map solutions to observationally equivalent solutions, and (defeasibly) taking such observational equivalence to consist in such solutions' representing the same instantiation of relative material distances and velocities, is not simply equivalent to committing the fallacy of inferential circularity, for the core *physical* criterion of observational equivalence itself is open to revision: a new theory (e.g., special relativity) might well end up suggesting a different criterion for observational equivalence (e.g., the same pattern of instantiation of material spatiotemporal *intervals*) from the theory that preceded it. There is nothing wrong with each theory suggesting its own criterion for observational equivalence; what is wrong is to assume that this criterion is not itself open to revision on the basis of theory.

6 The Invariance Principle

We now sketch how, in our view, a correct account of the **IP** should proceed. To keep things concrete, consider again the case of NGT and absolute velocity. NGT has an associated set $\mathcal{D}$ of DPMs. Now, we seem to have very good reason to think that, for instance, models related by a boost transformation are empirically indistinguishable: there is no possible measurement that any observer 'embedded' in such a Newtonian world could perform in order to determine what her absolute velocity actually is. Absolute velocity, in other words, is a quantity that no Newtonian observer could ever empirically *measure*. But why do we think this? It is sometimes claimed (see e.g. [15]) that this is a simple consequence of Newton's *laws*: that it follows merely from the fact that accelerations are left invariant under the Galilean transformations that absolute velocity is an undetectable physical quantity. As several authors (e.g. [1,4]) have pointed out, however, this is incorrect: in order to derive the conclusion that absolute velocity is a truly undetectable dynamical quantity one needs an additional assumption—one which in fact Newton implicitly made in his derivation of Corollary V in the *Principia*—namely, that both the inertial and gravitational mass of bodies (and the corresponding forces which act upon them) are *independent* of the bodies' absolute state of motion.[16] But this assumption is plausibly still only a necessary, and not a sufficient, condition to guarantee absolute velocity's undetectability. One additional (extremely obvious, but nevertheless non-trivial) assumption that needs to be made in order to ensure such undetectability concerns subjects' internal mental states: namely, that they do not covary in a systematic way with their absolute motion. That is, one must assume that subjects do not (e.g.) possess 'absolute velocity recorders' in the corners of their visual fields, or indeed any other devices which would allow them to have direct knowledge of what their own particular absolute velocity is at any given time.[17]

These assumptions are, of course, very natural ones to make. Indeed, most theorists writing on these topics appear to make them implicitly. What we wish to emphasise here, however, is simply their *non-triviality*; none of them, after all, follow purely from 'the truth' of Newtonian theory alone; if one of them were not to obtain, absolute velocity might well have turned out not to have been an undetectable quantity after all.

But why do these assumptions feel natural? One thing worth pointing out is that all of them seem to fit in with our 'scientific knowledge' in the broadest sense. We are, indeed, reasonably confident that forces do not systematically covary with systems' 'absolute' velocities (as our experiences on trains and aeroplanes, along with more detailed experiments, will testify); that humans do not, for instance, possess 'absolute

[16] Indeed, this assumption is required in order to ensure that NGT counts as a Galilean (or 'boost') invariant theory in the first place.

[17] This point in particular is emphasised by Roberts [27, §6]. As he notes, if such knowledge did exist it would be a very strange kind of knowledge indeed: in particular, it would not be communicable through standard physical channels. For instance, writing a letter to someone to inform him or her of what your absolute velocity actually is wouldn't work, as the relative positions of the ink particles on the written paper would be preserved in the boosted scenario as well.

velocity recorders' in the corners of their visual field; and that no other theory will eventually allow one to determine such a thing as one's 'true' motion with respect to some privileged inertial frame. It would thus seem to be the case that it is to *science itself*, in some very general sense, that we must appeal if we are to determine whether or not some specific quantity is detectable or not, and to which we must ultimately defer if we are to justify **IP**-based metaphysical inferences. In a certain sense, then, we should—as indicated above—embrace van Fraassen's *hermeneutic circle* [34, §3.7]: we should let science itself be our guide to what we (think we) can detect.

The proposal that we are suggesting here might be summarised as follows. First, one determines what the superfluous structure-indicating transformations on the solution space of one's theory actually are by engaging in hermeneutic circle-type reasoning: this allows one to determine which models of one's theory represent empirically indistinguishable, but nevertheless putatively physically distinct, solutions. *Our proposed epistemic approach to symmetries is then to define two models to be symmetry-related which bear this relation to one another.* Second, one justifies the excision of the superfluous 'variant' structure in question by appealing to a mixture of both scientific realism (i.e., the view that the models of one's theories should be construed more-or-less literally) and to Ockham's razor (i.e., the assumption that, other things equal, it is better for one's theory *not* to allow for physically distinct but empirically indistinguishable solutions).

Now, of course, in the Newtonian case things are fairly straightforward: working out which solutions are empirically indistinguishable is relatively easy once we have determined (by hermeneutic circle-type reasoning) that all that we in fact have empirical access to are the relative positions and velocities that material systems instantiate with respect to one another. *In more complex or heavily mathematicised theories, working out the empirical content of one's theory will almost certainly constitute a far less straightforward task.* Nevertheless, we think that the general process of reasoning that applies in these more complex cases will be the same in all essential respects to the one we are here suggesting applies in the simpler case of NGT. That is, in the first instance we engage in hermeneutic circle-type reasoning to work out which solutions in $\mathcal{D}$ represent empirically indistinguishable but physically distinct ways for the world to be; and in the second instance, we seek a novel, superior theory—or a superior interpretation or 'characterisation' of the original theory—according to which these two solutions are not only *empirically* indistinguishable, but represent *physically* identical states of affairs: in the case of the 'variant' quantity of absolute velocity in NGT, for instance, this will involve moving to Galilean (or, better, Newton-Cartan) spacetime.

This account constitutes an improvement over Ismael and van Fraassen's own proposal, insofar as it appears to solve what are arguably its most pressing difficulties. For here there is no distinction between the 'qualitative' and the 'measurable'; moreover, the account does not have the unwanted consequence that measurably distinct solutions can sometimes be indicative of the presence of superfluous structure; and furthermore the account would seem to be fully capable of accommodating the transition from Newtonian physics—in which we 'directly observe' spatial distances *simpliciter*—to relativity theory, in which we only 'directly observe' space*time* intervals. This is because on our account the hermeneutic circle serves (as it appears to have served for van Fraassen [34, §3.7]) an essential *dual* role: it is not only the means by which we determine *whether* two solutions are empirically indistinguishable; but it also, in a crucial and important sense, informs us of *what* it is that we think we can empirically detect.

The account also constitutes an improvement over those of Caulton and Dasgupta; in the former case, because more explicit detail regarding the "first phase of the interpretative process"—*viz.*, how two models of the theory in question are to be identified as empirically equivalent—is provided, and because we explicitly do not commit ourselves to the interpretational approach to symmetries; in the latter case, because the identification of two models as being empirically equivalent need not proceed in terms of notions of 'how things look', or 'observation sentences'.

7 Redundant Epistemic Symmetries

Let us now take a step back, and ask: it is, in fact, appropriate to take an epistemic approach to symmetry transformations at all? One significant concern regarding epistemic definitions of symmetries is the following: they constitute a very poor fit to scientific practice—in which symmetries are construed in a formal/ontic sense, prior to the introduction of notions such as observational equivalence.[18] Thus, while a philosopher might claim that (e.g.) Dasgupta's reducing in [10] symmetry-to-reality based reasoning to an Ockhamist norm to excise undetectable structure resolves the puzzle regarding why one should engage in symmetry-to-reality based reasoning, there is a sense in which this approach misses the mark—for the puzzle of why *physicists'* notions of symmetries (which are *not* stated in epistemic terms) should be involved in symmetry-to-reality based reasoning still remains.[19]

Our issue with the epistemic approach to symmetries, then, is that it appears to define itself out of a problem. The substantive question to be addressed is: "Why are symmetries a guide to reality?" The response on the part of the advocate of the epistemic approach is that we *define* symmetries in such a way that they *are* a guide to reality. But this is reminiscent of (e.g.) Strawson on induction [31, ch. 9] (Question: "Why is it rational to reason inductively?" Answer: "Because we defined it to be so."): giving this answer does not address the question of why symmetries—*as the term is used in physics*—are a guide to reality. To do this, one would need to (a) argue that such symmetries do indeed relate empirically equivalent models, and (b) appeal to an Ockhamist norm. Following this latter strategy (as we advocate) does not amount to defining symmetries in such a way that there is no interpretative question, and makes clear that the central task which must be addressed (on a case-by-case basis) is arguing *why* such-and-such symmetries do indeed relate empirically equivalent models.

Thus, the problems for epistemic approaches are not faced by formal/ontic definitions of symmetries. In this case, an antecedent notion of a symmetry transformation is presented; *if* one then discovers (not in this case by definition, but by some substantive argument—or perhaps the style of reasoning outlined in §6) that symmetry-related models are empirically equivalent, *then* one can apply the above-mentioned Ockhamist norm, to excise variant structure between symmetry-related models, in

[18]Our thanks to David Wallace for discussion on this point.

[19]We don't deny that physicists might have epistemic *motivations* for constructing their preferred definitions of symmetries. In that case, however, the puzzle would persist: how is it that said definitions of symmetry transformations, which are not epistemic in the sense given in this paper (even though they be *motivated* by considerations of detectability, etc.) do indeed come to be involved in symmetry-to-reality based reasoning?

line with the **IP**. In this case, one has discovered something substantial: one has discovered that symmetry-related models—where 'symmetry' is to be analysed in some prior terms—should be regarded as being physically equivalent (or at least, for the advocate of the motivational approach, one is *motivated* to regard symmetry-related models as being physically equivalent). By contrast, on epistemic conceptions of symmetries, one knows *from the outset* that symmetry-related models should be regarded as being physically equivalent (or, again, one is *motivated* to regard them as being physically equivalent), but *one has learnt nothing substantive about the theory in question by making this statement*.

In light of this, it should be clear that one *may* define a symmetry in general, epistemic terms, as outlined in the previous sections of this paper, but there is a certain sense in which *one gains nothing from doing so*. Where does this leave us? Insofar as one takes this point to be problematic, one might be inclined to return to ontic definitions of symmetries, such as those outlined by Belot [2, §§4-5]—the fact that in this case an antecedent definition of a symmetry transformation is given in terms of the formal structure of the theory in question means that the notion of a symmetry transformation is *not* rendered redundant, in the sense outlined above. Should the fact that there exists no *one*, univocal, ontic conception of a symmetry be construed as problematic? In our view, the answer is *no*—rather, there exist many different notions of symmetry transformations, each appropriate to different contexts, and some of which *may* be involved in symmetry-to-reality based reasoning, *if* a connection between symmetry and empirical equivalence can be forged on a case-by-case basis. Thus, *our own opinion on this matter is that one should embrace a plurality of ontic approaches to scientific theories*.[20]

Still, anyone proposing a return to ontic definitions of symmetries should tackle head-on Dasgupta's "problem of inferential circularity" [10, p. 864]. Recall that this

[20]One might point out that our own proposed epistemic account of symmetries is as open to the charge of redundancy as are other epistemic approaches to symmetries. However, to think that this thereby constitutes a serious problem with the paper's overall argument is, we think, to misunderstand the dialectic. To repeat, our purpose in this paper is to: (i) criticise extant epistemic approaches to symmetries; (ii) propose our own (improved) version of an epistemic approach; (iii) claim that *all* such approaches (*including* our own) are committed to a certain controversial claim (*viz.*, the conceptual redundancy of 'symmetry' and other related notions); and finally (iv) suggest that the obvious and natural way to avoid such a commitment (should one wish to avoid it!) is to adopt an alternative (e.g. ontic) approach to symmetries.

problem is the following: how is one to determine whether any such definition is correct—in the sense of constituting a genuine 'guide' to unreal structure—absent an antecedent understanding of which features of one's theories are truly physically real? Our response to this is twofold. First, one does not need to fix *all* physical structures in order to give an ontic definition of a symmetry—rather, one defines a symmetry transformation as that which preserves *some* physical feature F, then identifies undetectable structure which varies between (empirically equivalent) models which preserve F, then seeks a new formalism/interpretation in which that undetectable structure is excised. This requires only specification of *some*—not *all*—genuine physical features which must be preserved; and as long as this the case, ontic definitions of symmetries can be put to work—they are not viciously circular (although, one might say, they are still *partially* circular). Second, it is important to note (and, indeed, we have already done so above) that the identification of invariant physical features F between certain models of one's theories is *essentially tentative*, and relies on hermeneutic circle-style reasoning. That is, one's making use of an ontic definition of symmetries does not require a once-and-for-all demand that feature F be regarded as physical, and therefore preserved. Thus, contrary to the standard presentation of the problem of inferential circularity, deployment of ontic approaches to symmetry transformations does not rely upon an unassailable antecedent grasp of which features of our theories are physically real.

8 Close

Epistemic approaches to symmetry transformations are currently *à la mode*. In this paper, we have undertaken an in-depth critical analysis of three of the most prominent such accounts, due, respectively, to Ismael and van Fraassen (§3), Caulton (§4), and Dasgupta (§5). Building on what we take to be problems with those approaches, we have provided our own account of epistemic symmetries in §6. Having done so, we have raised in §7 the concern that epistemic approaches to symmetries render the general notion of a symmetry transformation redundant as a tool for metaphysical theorising about scientific theories; moreover, they render it at odds with physical practice. In light of this, we advocate pluralism about symmetry transformations: one should take it that there are many different possible definitions of symmetries, not all of which may be appropriate to every circumstance, but for each of which the discovery that symmetry-related models are empirically equivalent should constitute an important *post facto* discovery.

Acknowledgements

We are grateful to Adam Caulton, Sebastian de Haro, Abigail Holmes, Caspar Jacobs, Niels Martens, Nic Teh, David Wallace, and several anonymous referees, for valuable feedback on this work.

References

[1] Julian Barbour, *Absolute or Relative Motion? Vol. 1: The Discovery of Dynamics*, Cambridge: Cambridge University Press, 1989.

[2] Gordon Belot, "Symmetry and Equivalence", in R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics*, Oxford: Oxford University Press, pp. 318-339, 2013.

[3] Carolyn Brighouse, "Spacetime and Holes", PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, pp. 117-125, 1994.

[4] Harvey R. Brown, "Correspondence, Invariance, and Heuristics in the Emergence of Special Relativity", in S. French and H. Kamminga (eds.), *Correspondence, Invariance and Heuristics: Essays in Honour of Heinz Post*, Dordrecht: Springer Press, 1993.

[5] Jeremy Butterfield, "Against *Pointillisme* about Geometry", in F. Stadler and M. Stöltzner (eds.), *Proceedings of the 28th Ludwig Wittgenstein Symposium*, pp. 181-222, De Gruyter, 2006.

[6] Jeremy Butterfield, "Against *Pointillisme* about Mechanics", British Journal for the Philosophy of Science 57, pp. 709-753, 2006.

[7] Jeremy Butterfield, "Against *Pointillisme*: A Call to Arms", in D. Dieks, W. Gonzalez, S. Hartmann, T. Uebel and M. Weber (eds.), *Explanation, Prediction and Confirmation: New Trends and Old Ones Reconsidered*, pp. 347-365, Heidelberg: Springer, 2011.

[8] Adam Caulton, "The Role of Symmetry in the Interpretation of Physical Theories", Studies in History and Philosophy of Modern Physics 52, pp. 153-162, 2015.

[9] Shamik Dasgupta, "The Bare Necessities", Philosophical Perspectives 25, pp. 115-160, 2011.

[10] Shamik Dasgupta, "Symmetry as an Epistemic Notion (Twice Over)", British Journal for the Philosophy of Science 67(3), pp. 837-878, 2016.

[11] Talal Debs and Michael Redhead, *Objectivity, Invariance, and Convention*, Cambridge, MA: Harvard University Press, 2007.

[12] Neil Dewar, "Symmetries and the Philosophy of Language", Studies in the History and Philosophy of Modern Physics 52, pp. 317-327, 2015.

[13] John Earman, *World-Enough and Space-Time*, Cambridge, MA: MIT Press, 1989.

[14] Michael Esfeld and Vincent Lam, "Moderate Structural Realism About Space-Time", Synthese 160(1), pp. 27-46, 2008.

[15] Michael Friedman, *Foundations of Space-Time Theories*, Princeton, NJ: Princeton University Press, 1983.

[16] Carl Hoefer, "The Metaphysics of Space-Time Substantivalism", Journal of Philosophy 93(1), pp. 5-27, 1996.

[17] Jenann Ismael and Bas van Fraassen, "Symmetry as a Guide to Superfluous Theoretical Structure", in K. Brading and E. Castellani (eds.), *Symmetries in Physics: Philosophical Reflections*, Cambridge: Cambridge University Press, pp. 371-392, 2003.

[18] James Ladyman, "Scientific Structuralism: On the Identity and Diversity of Objects in a Structure", Aristotelian Society Supplementary Volume 81(1), pp. 23-43, 2007.

[19] James Ladyman and Don Ross (with David Spurrett and John Collier), *Every Thing Must Go: Metaphysics Naturalised*, Oxford: Oxford University Press, 2007.

[20] Thomas Møller-Nielsen, "Invariance, Interpretation, and Motivation", Philosophy of Science 84, pp. 1253-1264, 2018.

[21] Robert Nozick, *Invariances: The Structure of the Objective World*, Cambridge, MA: Harvard University Press, 2001.

[22] Oliver Pooley, *The Reality of Spacetime*, book manuscript, 2015.

[23] W. V. O. Quine, "On the Reasons for Indeterminacy of Translation", Journal of Philosophy 67(6), pp. 178-83, 1970.

[24] W. V. O. Quine, "On Empirically Equivalent Systems of the World", Erkenntnis 9(3), pp. 313-28, 1975.

[25] James Read, "Background Independence in Classical and Quantum Gravity", B.Phil. thesis, University of Oxford, 2016.

[26] James Read and Thomas Møller-Nielsen, "Motivating Dualities", Synthese 197, pp. 263-291, 2020.

[27] John T. Roberts, "A Puzzle about Laws, Symmetries, and Measurable Quantities", British Journal for the Philosophy of Science 59(2), pp. 143-168, 2008.

[28] Simon Saunders, "Physics and Leibniz's Principles", in K. Brading and E. Castellani (eds.), *Symmetries in Physics: Philosophical Reflections*, Cambridge: Cambridge University Press, 2003.

[29] Simon Saunders, "Mirroring as an A Priori Symmetry", Philosophy of Science 74(4), pp. 452-480, 2007.

[30] Howard Stein, "On Relativity Theory and Openness of the Future", Philosophy of Science 58(2), pp. 147-167, 1991.

[31] Peter Strawson, *Introduction to Logical Theory*, London: Methuen & Co. Ltd.,1952.

[32] Patrick Suppes, "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences", Synthese 12, 1960.

[33] Noel Swanson and Hans Halvorson, "On North's 'The Structure of Physics"', 2012. Available online at: http://philsci-archive.pitt.edu/9314/.

[34] Bas van Fraassen, *The Scientific Image*, Oxford: Clarendon Press, 1980.

[35] Bas van Fraassen, *Laws and Symmetry*, Oxford: Clarendon Press, 1989.