Synthese (2021) 199:315–344
https://doi.org/10.1007/s11229-020-02658-4

SYMMETRIES AND ASYMMETRIES IN PHYSICS

Sophistry about symmetries?

Niels C. M. Martens[1,2] · **James Read**[3]

Received: 16 September 2018 / Accepted: 8 April 2020 / Published online: 15 May 2020

Abstract

A common adage runs that, given a theory manifesting symmetries, the syntax of that theory should be modified in order to construct a new theory, from which symmetry-variant structure of the original theory has been excised. Call this strategy for explicating the underlying ontology of symmetry-related models *reduction*. Recently, Dewar has proposed an alternative to reduction as a means of articulating the ontology of symmetry-related models—what he calls (external) *sophistication*, in which the *semantics* of the original theory is modified, and symmetry-related models of that theory are treated *as if* they are isomorphic. In this paper, we undertake a critical evaluation of sophistication about symmetries—we find the programme underdeveloped in a number of regards. In addition, we clarify the interplay between sophistication about symmetries, and a separate debate to which Dewar has contributed—*viz.*, that between interpretational versus motivational approaches to symmetry transformations.

Keywords Symmetry · Symmetry-to-reality inferences · Reduction · Sophistication · Interpretational and motivational approaches

Contents

✉ Niels C. M. Martens
nmartens@uni-bonn.de; martens@physik.rwth-aachen.de

James Read
james.read@philosophy.ox.ac.uk

1 Lichtenberg Group for History and Philosophy of Physics, University of Bonn, Bonn, Germany

2 Institute for Theoretical Particle Physics and Cosmology, RWTH Aachen University, Aachen, Germany

3 Pembroke College, Univeristy of Oxford, Oxford, UK

1 Introduction

The venerable literature on symmetry transformations is brimming with new distinctions.[1] On the one hand is the question of whether symmetry-related models of a given theory should *invariably* be regarded as being physically equivalent, or whether the situation is more subtle. Advocates of the former view include e.g. Saunders (2003); for a response to this view, drawing upon the physics literature in order to advocate caution, see e.g. Belot's (2018). This former view was dubbed in Møller-Nielsen (2017) the *interpretational approach* to symmetries, as contrasted with a more modest *motivational approach* to symmetries, according to which symmetry-related models *at most motivate* one to construct a metaphysically perspicuous characterisation of the common ontology of symmetry-related models, but only once that characterisation is procured should those models be regarded as being physically equivalent.

A second distinction which has arisen in recent times in the literature on symmetry transformations regards the correct way to proceed in explicating the common ontology of symmetry-related models. (Note that this question is distinct from the above normative question regarding whether symmetry-related models may be regarded ab initio as being physically equivalent.) A common strategy (see e.g. Caulton 2015; Dirac 1930; Nozick 2001) states that, in order to so explicate the ontology underlying symmetry-related models, one should construct a new theory, trading only in the structures which are *invariant* across those symmetry-related models. Call this strategy *reduction*; this invariably involves modifying the *syntax* (i.e. the equations) of the original theory manifesting the symmetry under consideration. Recently, however, Dewar (2019) has proposed two alternatives to reductionism, which he dubs *internal sophistication* and *external sophistication* about symmetries. The former again involves mathematical reformulation,[2] this time not of the syntax but of the *semantics* (i.e., the models), in such a way that allows for the 'forgetting' of certain structures in the models of the original theory (what is meant by this will be made precise below).[3]

[1] See Brading and Teh (2017) for a recent survey of this literature, and Brading and Castellani (2003) for an older—though still exceptional—collection.

[2] Except for trivial cases in which the symmetry-related models are already isomorphic and one can immediately 'forget' certain structures—see Sect. 2.4.

[3] Dewar's way of articulating the syntax/semantics distinction does not coincide with another way of drawing that distinction: namely, syntax as the mathematics of one's theory (encompassing both syntax and semantics in Dewar's sense), and semantics as the *interpretation* of one's theory—i.e., the establishment of a mapping between the models of one's theory and possible worlds/physical situations. It is important to be clear on this difference, in order to avoid confusions going forward.

On the latter approach, by contrast, the semantics is modified without having first provided a mathematical reformulation: it is simply declared that symmetry-related models are treated *as if* they are isomorphic. Dewar claims in Dewar (2019) that external sophistication is not only easier to implement than reduction, but is also more general, in the sense that it is an interpretative strategy which is invariably available.[4]

In this paper, we undertake a critical evaluation of the sophisticationist strategy for articulating the ontology associated with symmetry-related models; we find in general this approach to be under-developed in a number of crucial respects. Moreover, while Dewar claims that sophisticated theories preserve the explanatory virtues of the original theories from which they are constructed, we argue that this claim is not invariably true. In addition, we seek to provide insight into the interplay between one's position in the interpretational/motivational debate and one's commitments in the reduction/sophistication debate.

The structure of this paper is as follows. In Sect. 2, we provide some essential background on the above-mentioned two debates in the philosophy of symmetries. In Sect. 3, we present the views of two notable figures who have contributed to these debates—namely, Dewar (see e.g. Dewar 2015, 2016, 2019) and Møller-Nielsen (see e.g. Møller-Nielsen 2015, 2017; Read and Møller-Nielsen 2020b, a)—; doing so will enable us to make clearer how the interpretation/motivation debate overlaps with the reduction/sophistication debate. In Sect. 4, we present three classes of criticism of sophistication about symmetries (as a justification of the interpretational approach). We conclude in Sect. 5.

2 Background

In this section, we present some essential groundwork on symmetry transformations in scientific theories. To be specific, in Sect. 2.1, we recall the 'semantic approach' to scientific theories—which will be the framework largely adopted in this paper. In Sect. 2.2, we present (following Dasgupta 2016) three approaches to the definition of symmetry transformations, and highlight the essential feature of symmetry transformations which will be relevant for our purposes in this paper. In Sect. 2.3, we recall the debate between 'interpretational' versus 'motivational' approaches to symmetry transformations, which regards the normative import of symmetries. Finally, in Sect. 2.4, we introduce the distinction which constitutes the focus of this paper—namely, that between 'sophistication' versus 'reduction' about symmetries. As we will see, this debate regards the best way to proceed in the construction and interpretation of one's physical theories, once one is presented with a theory manifesting certain symmetries.

[4] For other works engaging with distinctions very close to that between reduction and sophistication, see Rickles (2008), Sider (2018). Both Rickles (2008, ch. 8) and Dewar (2019, p. 514) highlight that, since symmetry reduction offers a path towards quantisation, it would be an interesting and worthy task to investigate the interactions between sophistication and quantisation; that task will have to wait for another day. For further discussion on symmetries and quantisation, see Belot (2003), Gryb and Thébault (2016).

2.1 Models

On the 'semantic conception' of scientific theories, 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 of tuples of specified geometrical objects.[5] Given a class $\mathcal{K}$ of KPMs for $\mathcal{T}$, one then restricts to the class of so-called 'dynamically possible models' (DPMs) $\mathcal{D} \subset \mathcal{K}$ of $\mathcal{T}$, by specifying certain dynamical equations which the geometrical objects in question must satisfy.

Let us illustrate this setup with an example. The KPMs of *Newtonian gravitation theory* (NGT) set in *Newtonian spacetime* are picked out by all tuples of the form $\langle M, t_{ab}, h^{ab}, \nabla, \sigma^a, \varphi, \rho, \xi^a \rangle$, where M is (as above) a four-dimensional differentiable manifold; t_{ab} and h^{ab} are fixed fields with respective signatures diag $(1, 0, 0, 0)$ and diag $(0, 1, 1, 1)$ and orthogonal in the sense that $t_{ab}h^{bc} = 0$; ∇ is compatible with t_{ab} and h^{ab}; σ^a is a fixed timelike (in the sense that $t_{ab}\sigma^b \neq 0$) and covariantly constant vector field representing the persisting points of absolute space; φ and ρ are real scalar fields on M representing, respectively, the gravitational potential and matter density field; and ξ^a is a timelike vector field, integral curves of which represent the motions of test particles.[6] Given the KPMs of NGT, the DPMs of this theory are picked out as those tuples $\langle M, t_{ab}, h^{ab}, \nabla, \sigma^a, \varphi, \rho, \xi^a \rangle$ which satisfy[7]

$$R^a_{bcd} = 0, \tag{2}$$

$$h^{ab}\nabla_a\nabla_b\varphi = 4\pi\rho, \tag{3}$$

$$-\nabla^a\varphi = \xi^b\nabla_b\xi^a. \tag{4}$$

Here, (2) imposes flatness of ∇; (3) is the *Newton-Poisson equation*, and (4) is Newton's force law in this context, where ξ^a represents the four-velocity of the test particle under consideration.

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 equivalence classes of such models. In the case in which the interpretation of $\mathcal{T}$ leads to such a

[5] Here, we understand 'geometrical object' in the sense of Anderson (1967); for an alternative understanding of the meaning of this term, see Martens and Lehmkuhl (2019).

[6] For further details on the mathematical structure of the KPMs of NGT, see Malament (2012), Pooley (2015).

[7] An anonymous referee has objected that this presentation of NGT is ill-formed, in the sense that ρ does not obey its own dynamical equations. To this, we would reply that one should distinguish (i) an object in a theory satisfying its own dynamical equations, from (ii) an object in a theory being permitted to vary from model to model. (ii) may hold of a given object (e.g. ρ) without (i).
Even if one does wish to afford ρ its own dynamics, though, there are many moves that one could make. For example, one could take the theory to be a theory of dust, in which case ξ^a would be added to the theory's defining tuple as the velocity field of the dust and the system of equations would be supplemented by a conservation law,

$$\nabla_a\left(\xi^a\rho\right) = 0. \tag{1}$$

We are grateful to the anonymous referee for this suggestion.

redundancy, we may attempt to construct a reduced space of models $\tilde{\mathcal{D}}$ of some new theory $\tilde{\mathcal{T}}$, in which equivalent models of $\mathcal{T}$ are 'mathematically identified'—in the sense that a formal mapping is established between the equivalence classes of DPMs of $\mathcal{T}$ and *unique* DPMs of $\tilde{\mathcal{T}}$. We will see concrete examples of these manoeuvres in the ensuing sections of this paper.

2.2 Symmetry transformations

There is a rich philosophical literature on the definition of symmetry transformations in physics. A useful tripartite distinction between *formal*, *ontic*, and *epistemic* approaches to symmetry transformations is drawn by Dasgupta (2016, §5). According to *formal* definitions of symmetries, a symmetry is an automorphism of the space of DPMs of the theory in question, preserving some specified formal property. One trivial example of a formal definition of a symmetry—presented in Dasgupta (2016, §5.2) and critiqued compellingly in Belot (2013, p. 6)—is that a symmetry is any transformation which preserves the dynamical equations of the theory in question. (Formal definitions of symmetries face a natural charge of physical irrelevance; for further discussion, see Dasgupta 2016; Read and Møller-Nielsen 2020b.) According to *ontic* definitions of symmetries, a symmetry transformation is an automorphism of the space of DPMs of a given theory, preserving some specified class of putative *physical* quantities. (Examples of ontic definitions of symmetries include e.g. Lagrangian symmetries, generalised Noether symmetries, etc.—see Belot (2013).[8]) Finally, according to *epistemic* definitions of symmetries, a symmetry transformation is an automorphism of the space of DPMs of a given theory, such that any two models related by that mapping are empirically equivalent. (In this sense, symmetry-related models of necessity agree on 'empirical substructures'—see Van Fraassen 1980, p. 64.)

Note that it may or may not be the case that models of a given theory related by formal or ontic symmetries are empirically equivalent. However, for the purposes of this paper—and as will be explained in detail in the following sections—the symmetry transformations of interest are precisely those which (whether by definition or otherwise) *are* regarded as relating empirically equivalent models. Thus, we make this restriction in the remainder of this paper, while remaining officially neutral on the most appropriate *definition* of a symmetry transformation (for more on this latter topic, see Read and Møller-Nielsen 2020b).[9]

2.3 Interpretation and motivation

The above is purely formal; there remains an outstanding question concerning *when* two models of $\mathcal{T}$ should be interpreted as representing the same possible world. One popular line is what was dubbed in Møller-Nielsen (2017, §2) the *interpretational* approach to symmetry transformations: two symmetry-related models of $\mathcal{T}$ typically may be regarded ab initio as representing the same possible world, even in the absence

[8] One would be right to question whether the distinction between ontic and formal definitions of symmetries is clear-cut—cf. (Read and Møller-Nielsen 2020b, §2.2).

[9] In making this point, we echo Ismael and van Fraassen (2003).

of a coherent explication of their common ontology.[10] (For an extensive list of citations of authors embracing this line, see (Read and Møller-Nielsen 2020a, §3.1).) This is in contrast with the *motivational* approach to symmetry transformations (Møller-Nielsen 2017, p. 4), according to which the existence of symmetry-related models *at most motivates* us to provide an explication of the shared ontology of these symmetry-related models, but only once such an explication is provided is it legitimate to regard those models as representing the same possible world. (For detailed discussion of the interpretation/motivation distinction, see Møller-Nielsen 2017; Read and Møller-Nielsen 2020a, b.)

It is worth highlighting that there exist two different strands of interpretationalism—the issue here regards the 'typically' clause in the above formulation of the position. On the stronger version of the interpretational approach, this clause is redundant: symmetry-related models may *invariably* be regarded ab initio as being physically equivalent. On the weaker version of the interpretational approach, by contrast, this 'typically' clause is *not* redundant, and invites a certain hedging: advocates of this weaker version of the view may argue that, in virtue of certain e.g. theoretical/metaphysical/super-empirical considerations, symmetry-related models should not *invariably* be regarded ab initio as being physically equivalent—for it may be that in certain cases, one has strong independent reasons to continue to regard these models as being distinct.[11] We will see in Sects. 3.1 and 4.2 that Dewar should be understood as falling into the latter of these two camps.

Once this distinction is noted, one may also carry it across to the motivational approach to symmetries. Specifically, on the weaker version of the motivational approach, symmetry-related models may only be regarded as being physically equivalent once we have to hand a coherent metaphysical picture of the common ontology underpinning their equivalence—but once we are in possession of that picture, other e.g. explanatory/metaphysical factors should not bear upon our regarding those models as being physically equivalent. By contrast, on the stronger version of the motivational approach, symmetry-related models may only be regarded as being physically equivalent once we have to hand a coherent metaphysical picture of the common ontology underpinning their equivalence—*and*, moreover, even once we are in possession of such a picture, it may be that certain explanatory/metaphysical considerations preclude us from regarding those models as being physically equivalent.

Where do specific authors stand with respect to these distinctions? Very briefly, we take Saunders (2003) to subscribe to strong interpretationalism; Dewar (2015, 2019) to weak interpretationalism (see Sects. 3.1 and 4.2); Møller-Nielsen (Møller-Nielsen

[10] This 'coherent explication' is what Møller-Nielsen calls in Møller-Nielsen (2017) a 'metaphysically perspicuous characterisation' of the common ontology of symmetry-related models. Though we discuss this notion below, here is one example of such a characterisation: Galilean spacetime (in which the vector field σ^a of Newtonian spacetime is excised) affords the metaphysically perspicuous characterisation of the common ontology underlying models of NGT set in Newtonian spacetime related by kinematic shifts (which are defined with respect to the σ^a field). (Cf. e.g. Earman 1989, ch. 3.)

[11] Suppose, for example, that one's metaphysics is built around the notion of undetectable absolute velocity—then, even an interpretationalist may wish to resist regarding models of Newtonian gravitation theory set in Newtonian spacetime as being physically equivalent.

2017; Read and Møller-Nielsen 2020a) to weak motivationalism;[12] and Martens (2018b) to strong motivationalism. While these points are important, however, it is worth re-emphasising the central difference between the interpretational and motivational approaches. Interpretationalists *of all stripes* maintain that one's regarding symmetry-related models of a given theory as being physically equivalent *need not wait upon an explication of the common ontology thought to underpin that physical equivalence*. It is *this* point which, crucially, separates the interpretationalist from the motivationalist.

2.4 Reduction and sophistication

The central focus of this paper falls upon what Dewar dubs a distinction between *reduction* versus *sophistication* about symmetry transformations. Here is how he puts the matter:

> It is often claimed that the symmetries of a theory reveal "surplus structure": structure which, in some sense, the theory could do without. For example, the boost symmetry of Newtonian mechanics indicates the superfluousness of absolute velocities; the gauge symmetry of electromagnetism reveals the superfluousness of absolute potentials; and so on and so forth. Moreover, it is widely held that if this is the case, then some modification of one's theory is appropriate, so as to make explicit what structure is *not* surplus (e.g. the replacement of Newtonian by Galilean spacetime, in response to the boost symmetry of Newtonian mechanics). ... I compare and contrast two ways of making such a modification. The first is to replace the theory by (what I shall call) a *reduced* theory: a theory that deals only in quantities which are invariant under the relevant symmetry. The second is to replace the theory by (what I shall call) a *sophisticated* theory: a theory in which models related by a symmetry are isomorphic. (Dewar 2019, pp. 485–486)

Reductionism—i.e., the advocation of the construction of such a 'reduced' theory when confronted with symmetry-related models of one's original theory in order to explicate the common ontology of those models—is certainly a widespread view in the literature—for presentations of such a view, see e.g. Butterfield (2018), Caulton (2015), De Haro and Butterfield (2018).[13] In more detail,

> the idea [of reductionism] is that we (i) identify some collection of invariants of the original theory; (ii) specify a theory in terms of those invariants; and (iii) show that the new theory captures all the symmetry-invariant content of the old theory. (Dewar 2019, pp. 492–493)

[12] More precisely, Møller-Nielsen's formulation of his preferred motivational approach equivocates between the weak and the strong version. Martens has argued that Møller-Nielsen's analysis of electromagnetism indicates that what he has in mind is the weak version (Martens 2018b). In fact, according to Martens, this case study is one of the main reasons that one should favour the strong version instead.

[13] We take the call of De Haro and Butterfield to find a 'common core' in the presence of symmetry-related models, or duality-related theories, to manifest the reductionist view. (We do not discuss further 'dualities' in this paper; for recent work on this topic, see e.g. Butterfield 2018; De Haro and Butterfield 2018; Matsubara 2013; Polchinski 2017; Read 2016; Rickles 2011.)

To put the matter differently, we take reductionism to consist in the following. Take the space of solutions $\mathcal{D}$ of the theory under consideration, and consider all classes of symmetry-related models in $\mathcal{D}$ (where, as always, we restrict to the relevant class of symmetry-related models which are empirically equivalent—cf. Sect. 2.2). Then, construct the space of DPMs $\tilde{\mathcal{D}}$ of some new theory,[14] such that the classes of symmetry-related models in $\mathcal{D}$ are mapped to a unique element of $\tilde{\mathcal{D}}$, which contains the 'common mathematical structure' of the original class of elements of $\mathcal{D}$. This fulfills what Dewar calls the mandate to "specify a theory in terms of [the] invariants" of the original theory; articulating the mapping between the classes of symmetry-related elements of $\mathcal{D}$ and the unique elements of $\tilde{\mathcal{D}}$ suffices to "show that the new theory captures all the symmetry-invariant content of the old theory." Reductionism, then, involves a certain kind of mathematical identification in the sense of Sect. 2.1—namely, mathematical identification in which the 'reduced' theory traffics only in "invariants of the original theory."

The alternative to reductionism is what Dewar calls *sophistication* about symmetries. Here,

> the idea is that we need not insist on finding a theory whose models are invariant under the application of the symmetry transformation [as in the case of reductionism], but can rest content with a theory whose models are isomorphic under that transformation. That is, if M and N are symmetry-related models of the unreduced theory, then they give rise to the same model of the reduced theory ... ; the proposal is that we instead look for a theory such that M and N give rise to distinct but isomorphic models. (Dewar 2019, p. 498)

If one actually constructs a new theory, in which "M and N give rise to distinct but isomorphic models", and which is related to the original theory by some 'forgetful' map (see (Dewar 2019, p. 502) for details, and e.g. (Weatherall 2016) for further discussion), then one has (in Dewar's terminology) *internally sophisticated* one's theory.

At (2019, p. 502), Dewar offers the following remark: the difference between reduction and sophistication is, essentially, that while the former position advocates altering the *syntax* (i.e. equations) of the theory under consideration in the presence of symmetry-related models in order to articulate the common content of those models, the latter position advocates altering the *semantics* (i.e. models) of that theory, such that "the pictures on the new semantics are simply what we obtain by taking the old objects, and *declaring*, by fiat, that the symmetry transformations are now going to "count" as isomorphisms". While this is certainly true in the case of external sophistication (on which more below), it is worth registering that (just as in the case of reduction) this claim is not entirely true in the case of internal sophistication—for here one is to *reformulate* the original theory (i.e., modify the semantics of the original theory) such that the interpretation now proceeds in terms of the 'naïve' interpretation of the models of the *new* theory—*where that new theory was constructed by modifying the*

[14] Clearly, this will involve first constructing a space of KPMs $\tilde{\mathcal{K}}$ for that new theory.

semantics (i.e., the models) of the original theory in a way that enables the possibility of 'forgetting' structure.[15,16]

To get clearer on what *external* sophistication is supposed to be, it will first be useful to distinguish it from what one might call *traditional sophistication*. Of sophistication in general, Dewar states explicitly that

> the proposal on the table—that we can do justice to a symmetry using isomorphism rather than invariance—is a generalisation of the "sophisticated substantivalist" method for dealing with spacetime symmetries. (Dewar 2019, p. 501)

Recall that 'sophisticated substantivalism' affords a means of regarding hole-diffeomorphic models of general relativity (GR),[17] or statically-shifted models of NGT, as being physically equivalent—it does so by rejecting the view that spacetime points have primitive identities which persist across possibilities. (See Pooley 2002, 2015 for details.) Dewar borrows the term 'sophistication' from these debates, but admits that the notion of sophistication has been loosened substantially in his hands, as the original concept is associated with an attitude *only* towards symmetry-related models that *are already isomorphic*. Specifically, sophisticated substantivalism is a metaphysical thesis, regarding how to interpret the ontology of isomorphic symmetry-related models.[18] It is this view which we call *traditional sophistication*.

Dewar's *external sophistication*, on the other hand, is the statement that symmetry-related models should be regarded as being isomorphic, with (in general) no (explicit) accompanying metaphysical package.[19] This is the attitude which Dewar advocates in cases in which the symmetry-related models under consideration are *not* isomorphic. In this case, the procedure is more complicated, for (i) the models under consideration must be interpreted 'as if' they were isomorphic (which is now non-trivial, since they actually are not); and (ii), the traditional sophisticationist methodology must be applied in order to regard those models—interpreted as being isomorphic—as in fact representing the same physical states of affairs. We will see in Sect. 4 that the absence of a suitably robust metaphysical package accompanying (i) constitutes our central concern with external sophistication: we fail to understand how the above 'as if' can do the metaphysical work required of it.

[15] The examples presented in Weatherall (2016) provide a clear illustration of the differences between reduction and internal sophistication. For instance, in the case of electromagnetism formulated in terms of the vector potential A^a, the reduced version of this theory (where the reduction proceeds with respect to the $U(1)$ gauge symmetry of the theory) is electromagnetism formulated in terms of the Faraday tensor F_{ab}, while the internally sophisticated version of the theory is the fibre bundle formulation of electromagnetism. More on this in Sect. 4.

[16] Dewar's, in this sense, 'putting dynamics before kinematics', invites comparisons with Brown's *dynamical approach* to physical theories (Brown 2005). Though interesting, we defer these comparisons to future work.

[17] For background on the hole argument, see e.g. Norton (2019).

[18] For a generalisation of sophisticated substantivalism to the non-spatiotemporal case, see Esfeld and Lam's 'moderate structuralism' (Esfeld and Lam 2011), and our discussion below.

[19] That said, it is worth flagging that Dewar takes it that his view *does* have metaphysical content—this will be discussed in depth below.

At this point, it might be helpful to present two alternative plans which sophisticationists seem to seek to adopt. The first, for internal sophisticationists, runs as follows:

1. *Semantically reformulate:* Identify an alternative mathematical formalism for the theory in question by modifying the semantics (*nota bene*: not the syntax, or this would be a reduced formalism) such that the models corresponding to symmetry-related models in the original theory are isomorphic.
2. *Traditionally sophisticate:* Apply the sophisticated substantivalist methodology (or some suitable analogue—see below) in order to regard these isomorphic models of the new formalism as representing the same physical state of affairs.

The second plan, this time for external sophisticationists, is the following:

1. *Declare isomorphic:* Declare that the symmetry-related models under consideration are to be treated 'as if' they are isomorphic.
2. *Traditionally sophisticate:* Apply the sophisticated substantivalist methodology (or some suitable analogue—see below) in order to regard these isomorphic models of the new formalism as representing the same physical state of affairs.

Here, external sophisticationism can be understood as explicitly skipping the task of semantic reformulation, undertaken by the internal sophisticationist: the ontology of symmetry-related models can be articulated by treating them 'as if' they are isomorphic, then (at least implicitly) recoursing to traditional sophistication. By contrast, internal sophisticationism wishes to realise the sophisticationist strategy by explicitly constructing the appropriately mathematically reformulated theory in terms of which the ontology of the models of the original theory is to be understood—except for trivial cases in which the symmetry-related models are already isomorphic and one can immediately traditionally sophisticate. It is thus important to note that internal sophistication is understood as an alternative to, rather than a special case of, external sophistication. (This will be especially relevant in Sect. 4.1, where we argue that several of Dewar's examples of sophistication, being examples of internal or even merely traditional sophistication, do therefore not support the universal applicability of external sophistication.)

In sum, then, our taxonomy of 'strains of sophistication' is the following:

> *Traditional sophistication:* The decision to be anti-haecceitist or anti-quidditist (as appropriate—see below) about a theory in which symmetry-related models are already isomorphic.
> *Internal sophistication:* The view that if symmetry-related models are not isomorphic, then one should seek an explicit semantic mathematical reformulation which renders them isomorphic, and then interpret those models in an anti-haecceitistic/anti-quidditistic fashion.
> *External sophistication:* The view that if symmetry-related models are not isomorphic, one can simply *declare* them isomorphic (without seeking an explicit reformulation), and then apply the appropriate anti-haecceitist or anti-quidditist interpretation.

To close this subsection, it is worth noting that Dewar's external sophistication is very much akin to a view which Sider calls *quotienting* (Sider 2018, ch. 5)—indeed, in this paper we take these to be the very same view.[20] As Sider puts it, this is the view

> according to which, roughly, we can say *that* theories are equivalent without saying *why* they are equivalent in terms of fundamentality and underlying third theories.[21] (Sider 2018, p. 152)

Sider continues that, according to the quotienting perspective,

> [t]here may be no way to say what is "really" going on; maybe every good model has artifacts. It's ok to just say: this model does a good job of representing the phenomenon, but certain features of the model are artifacts. Moreover, for any model, we can say which features of the model are genuinely representational and which are artifacts. There is no need to provide some privileged, artifact-free description from which we can recover this information. (Sider 2018, p. 153)

With this view on the table, we will shortly turn to a critical appraisal of external sophistication—we will end up agreeing with Sider that this is *not* a viable approach to articulating the ontology of symmetry-related models. Before doing so, however, we must get clear on how Dewar's views on all of the above distinctions interplay with one another; it will also be informative to consider the positions of certain other relevant authors on these matters.

3 Notable positions

In this section, we present and discuss the views of Dewar (Sect. 3.1) and Møller-Nielsen (Sect. 3.2) on the interplay between the debates on interpretation/motivation and on reduction/sophistication.

3.1 Dewar's views

Dewar is an avowed interpretationalist about symmetries—this is evident when he writes at (2015, p. 317) that

[20] Clearly, nomenclature here is not optimal, for one might take it that 'to quotient' is synonymous with 'to reduce'. The reader is cautioned not to conflate reduction and quotienting—for the latter is the same as (external) sophistication!

[21] Compare this with the following quote from an earlier paper by Dewar, which presages the external sophisticationist position:

> A more interesting thought, then, would be to ask whether there is some way in which we could be anti-realist about part of a model without being required to explicitly single out the parts of the model one is anti-realist about. I think the answer is yes. The trick is to stipulate which models are synonymous, rather than specifying which bits of a model one reads literally or not: we express our qualified-realist attitude by affirming certain non-isomorphic models as synonymous, which commits us to denying that the respects in which such models disagree correspond to any physically significant difference. (Dewar 2015, p. 322)

> It is the contention of this paper that models related by a symmetry transformation are merely different ways of representing the same physical state of affairs ...

The same thought is evident from the very first line of his D.Phil. thesis (Dewar 2016, p. 3):[22]

> This thesis examines the idea that when a physical theory contains symmetries, the theory should be interpreted in such a way that symmetry-related models represent the same physical state of affairs.[23]

How does Dewar think that a defence of the interpretational approach is supposed to go? Prima facie, such a view faces the obvious difficulties: (Cf. Møller-Nielsen 2017; Read and Møller-Nielsen 2020a.)

1. How are we to identify the common structure associated with symmetry-related models—and have we any reason to think that such structure is always there to be found?
2. Even supposing that such structure can be found, does it invariably admit of a coherent physical interpretation?
3. Even if such an interpretation is available, does it satisfy all super-empirical criteria that one may consider relevant?[24]

In fact, Dewar is aware of and sympathetic to these issues. For instance, on (1) and (2), he writes at (2019, p. 495) that

> it is highly non-trivial to find such a reduced theory—or even to demonstrate with confidence that such a theory could exist.

On (3), Dewar acknowledges at (2019, p. 496) that

> even if such a theory can be found, that theory may seem to have explanatory deficits relative to the original theory.

[22] Compare this to the following quote by Dewar:

> In this article, I will suppose that, at least under certain circumstances and for certain theories, the following claim is true:
>
> > For a theory containing symmetries, we should not interpret that theory in such a way that the symmetry-related models (that is, models related by a map induced by a symmetry) represent distinct ways for the world to be. (Dewar 2019, p. 491)

These restrictions to certain circumstances and theories, which are not further specified by Dewar, may seem inconsistent with the universal scope of both quotes in the main text, unless the restrictions only concern super-empirical criteria (i.e. weak interpretationalism). Further restrictions would give the game away to motivationalism, since the main point of contention between interpretationalism and motivationalism (bracketing the super-empirical considerations, Sect. 2.3) is exactly universalism, i.e. whether one can say for all symmetries that symmetry-related models represent the same possible state of affairs (interpretationalism) or merely for some symmetries—namely those for which one has explicitly provided a certain mathematical reformulation (motivationalism). We will thus assume that the restrictions concern at most super-empirical criteria.

[23] One might reasonably note that Dewar says 'examines', rather than 'endorses'. That the latter is true is, however, unequivocal for the reader of the piece.

[24] As will become clear below, while we recognise internal sophistication, like reduction, as being a legitimate means of articulating the common ontology of symmetry-related models, this approach also faces difficulties (1)–(3).

In light of these problems, Dewar asks, "Is there some other way of taking on board the above interpretational principle [*viz*., the interpretational approach to symmetries], without seeking out a reduced theory?" (Dewar 2019, p. 498). It is at this juncture that sophistication enters the picture. The thought is that (external) sophistication itself *typically* affords a coherent explication of the ontology of the symmetry-related models under consideration—thereby justifying the interpretational approach. Call this *universal external sophistication*.

Once again, the 'typically' qualifier may be read either in a strong sense—'invariably'—or in a weak sense, if one allows that extra-empirical considerations may block the sophistication (Sect. 2.3). Note that Dewar's speaking of 'explanatory deficits' in response to (3), as well as his discussion of several case studies, some of which we will revisit in Sect. 4.2, seem to place him in the camp of the weaker strand of interpretationalism. On the other hand, he also believes that the tentative explanatory powers of an unreduced, unsophisticated theory are often ultimately dispensable (Dewar 2015) and seems to claim—intending to improve upon reductionism—that (external) sophistication will *invariably* preserve explanations that are indispensable (Dewar 2019) (more on this in Sect. 4.2). This would render the difference between the weak and strong forms of the interpretational approach moot, at least in practice. The weak form effectively becomes the strong form, at least with respect to explanatory considerations, as it will never in fact occur that these considerations block sophistication.

It is precisely the promise of universal external sophisticationalism in support of (weak) interpretationalism that we call into question in Sect. 4 of this paper. We contend that external sophistication *in itself* does not afford a perspicuous explication of the ontology of symmetry-related models (Sect. 4.3). One *can* sometimes obtain such a perspicuous explication, but only when sophistication is accompanied by mathematical reformulation and appropriate interpretation of that new formalism (this, to repeat, being what Dewar calls 'internal sophistication'). Since it is however not clear, a priori, whether such a reformulation can *invariably* be found, one has to explicitly provide one for each symmetry in each theory. But this is just to combine (internal) sophisticationism with motivationalism about symmetries, not with interpretationalism (i.e. option (e) in Sect. 3.2 instead of option (c)) (Sect. 4.1). We will moreover question the claim that sophistication will *invariably* preserve explanations (Sect. 4.2).

3.2 Møller-Nielsen's views

It is illuminating to contrast the views of Dewar on the debates between interpretational and motivational approaches to symmetries, and between sophisticationism and reductionism about symmetries, with those of another author who has written on these matters—namely, Møller-Nielsen (Møller-Nielsen 2015, 2017; Read and Møller-Nielsen 2020a, b). While Møller-Nielsen explicitly favours the motivational approach to symmetries (indeed, the principal aim of Møller-Nielsen 2017; Read and Møller-Nielsen 2020b is to defend this approach), what is more interesting is that, depending upon the case in question, he favours reduction over sophistication, or

vice versa, as a means of providing a perspicuous characterisation of the ontology of symmetry-related models.

Before spelling this out in detail, it will be useful to clarify a related issue which has arisen in the recent literature: the appropriate notion of 'isomorphism'. To this end, consider a model of NGT set in Newtonian spacetime, $\mathcal{M} = \langle M, t_a, h^{ab}, \nabla, \sigma^a, \varphi, \rho, \xi^a \rangle$, and the following further two models of this theory:

1. A *statically-shifted* model $\mathcal{M}_{\text{stat}} = \langle M, t_a, h^{ab}, \nabla, \sigma^a, \alpha_* \varphi, \alpha_* \rho, \alpha_* \xi^a \rangle$, where α implements a static shift (i.e., time-independent translation) of the material content of the universe.
2. A *kinematically-shifted* model $\mathcal{M}_{\text{kin}} = \langle M, t_a, h^{ab}, \nabla, \sigma^a, \beta_* \varphi, \beta_* \rho, \beta_* \xi^a \rangle$, where β implements a kinematic shift (i.e., linearly time-dependent transformation) of the material content of the universe.

At (2017, p. 1260), it is stated by Møller-Nielsen that $\mathcal{M}$ and $\mathcal{M}_{\text{stat}}$ are isomorphic, for "they represent worlds that differ at most with regard to which particular objects are playing which qualitative roles (i.e., they represent at most haecceitistically distinct possible worlds)" (Cf. Pooley 2015, p. 70 and Read 2016, p. 221). It should be clear, on reflection, that this is *not* an appropriate definition of 'isomorphism', for it involves *interpretation*—but isomorphism should be a formal, mathematical notion.

To make clear the relations between these two notions of isomorphism (i.e., the above interpretative notion on the one hand, and the formal, mathematical notion on the other), in the context of spacetime theories, consider, following Weatherall (2018), certain maps which one could define between elements of the solution space of NGT. In particular, consider the following two maps:

- $1_M : M \to M$, which is the identity map on M. This is the unique map such that, given any other map $\gamma : M \to M$, $\gamma \circ 1_M = 1_M \circ \gamma = \gamma$.
- $\psi : M \to M$, which is a diffeomorphism (i.e., a smooth map with smooth inverse) taking all geometrical objects O on M to the pushforward geometrical object, $\psi_* O$.

Now, as Weatherall stresses, "according to the theory of smooth manifolds, diffeomorphism is the standard of isomorphism for manifolds; just as other mathematical objects are only defined up to isomorphism, manifolds are only defined up to diffeomorphism" (Weatherall 2018, p. 335). In the case of the static shift, $\mathcal{M}$ and $\mathcal{M}_{\text{stat}}$ are *not* identical mathematical objects, for they are are not related by 1_M; however, they *are* isomorphic as manifolds, for α implements a symmetry of Newtonian spacetime, i.e., $\langle M, t_a, h^{ab}, \nabla, \sigma^a \rangle = \langle M, \alpha_* t_a, \alpha_* h^{ab}, \alpha_* \nabla, \alpha_* \sigma^a \rangle$, in which case one may write $\mathcal{M}_{\text{stat}} = \langle M, \alpha_* t_a, \alpha_* h^{ab}, \alpha_* \nabla, \alpha_* \sigma^a, \alpha_* \varphi, \alpha_* \rho, \alpha_* \xi^a \rangle$, illustrating that $\mathcal{M}$ and $\mathcal{M}_{\text{stat}}$ are related by a diffeomorphism α. By contrast, in the case of the kinematic shift, $\mathcal{M}$ and $\mathcal{M}_{\text{kin}}$ are (again) not identical mathematical objects, for they are are not related by 1_M; moreover, they are are also not isomorphic as manifolds, for β does not implement a symmetry of Newtonian spacetime, for $\sigma^a \neq \beta_* \sigma^a$, meaning that $\mathcal{M}$ and $\mathcal{M}_{\text{kin}}$ are not related by a diffeomorphism β.

All of this amounts to the following. While the 'interpretative' definition of isomorphism (favoured by Møller-Nielsen) issues the right verdict on the static and kinematic shifts, it does so for the wrong reasons: it is (in our view) preferable to use the mathematical definition of isomorphism, and *only then* invoke interpretative notions. In

the above case: having witnessed that models of NGT related by a static shift are isomorphic (in the mathematical sense), whereas models related by a kinematic shift are not, one can then maintain (as is, by now, standard in the literature) that, in order to undercut the possibility of a static shift, isomorphic models should be interpreted anti-haecceitistically (and no mathematical reformulation is necessary in order to do this—see Møller-Nielsen (2017, p. 1260) and Pooley (2015, p. 229)); by contrast, in order to undercut the possibility of a kinematic shift, mathematical reformulation is necessary, in order to interpret those models as corresponding to isomorphic models of some *new* theory, and then invoke anti-haecceitism or anti-quidditism.

With the above clarifications in hand, let us return to considering Møller-Nielsen's views in the two debates under consideration. When symmetry-related models are *not* isomorphic (in his sense—but note that this is extensionally equivalent to the mathematical notion of isomorphism in the cases under consideration here), as for e.g. $\mathcal{M}$ and $\mathcal{M}_{\text{kin}}$, Møller-Nielsen favours reduction as a route towards explicating the common ontology of those symmetry-related models: the thought is that only by constructing such reduced models and then interpreting them can such an explication be procured (we concur with this verdict, modulo the possibility of internal sophistication as a distinct means of articulating the common ontology of these symmetry-related models;[25] more on this in Sect. 4 below).[26]

By contrast, in cases in which symmetry-related models *are* isomorphic, Møller-Nielsen does not favour reductionism, but rather (traditional) sophistication. For example, in the case of solutions of general relativity related by a hole diffeomorphism, or of solutions of NGT set in Newtonian spacetime related by a static shift (i.e., $\mathcal{M}$ versus $\mathcal{M}_{\text{stat}}$) Møller-Nielsen claims that sophisticated substantivalism (as presented above) affords a perspicuous characterisation of the common ontology of these models (namely, a characterisation proceeding on the basis of anti-haecceitism)—no reductionist move proceeding in terms of e.g. Einstein algebras or Leibniz algebras (respectively—cf. Earman 1989, ch. 9) is necessary. It is clear, then, that while Møller-Nielsen embraces traditional sophistication, he rejects external sophistication; as a consequence, he appeals to reductionism as a means of explicating the ontology of non-isomorphic, symmetry-related models, but sophisticated substantivalism as a means of explicating the ontology of isomorphic, symmetry-related models.

What is our purpose in making these observations? The reason for doing so is the following: the case of Møller-Nielsen illustrates that the interpretational/motivational

[25] But see footnote 24.

[26] An anonymous referee has questioned whether the move from NGT set in Newtonian spacetime, to NGT set in Galilean spacetime, is best understood as a case of reduction, or of internal sophistication. Insofar as one simply excises ('forgets') σ^a from the models of the former theory, we agree that this move is, in fact, best understood as a case of internal sophistication. There are, however, significant subtleties here, for in fact (see Read and Møller-Nielsen 2020b, fn. 9), following Pooley (2015, §§4.4–4.5), Møller-Nielsen does *not* favour the formulation of NGT set in Newtonian spacetime presented above; rather, he makes use of models of this theory which do not feature the derivative operator ∇ (since, in fact, σ^a also provides a standard of straightness of paths), and favours dynamics for the theory written directly in terms of σ^a (rather than ∇). In that case, the move from NGT set in Newtonian spacetime to NGT set in Galilean spacetime also involves modification of the equations of the theory—i.e., the *syntax*—in order to arrive at the usual formulation of the laws of the latter theory (i.e., (2)–(4)), formulated using the derivative operator ∇. Given this, it is arguably best to understand this change of spacetime setting for NGT, for Møller-Nielsen, as a case of *reduction*.

distinction is prima facie orthogonal to the reduction/sophistication distinction—it might be that:

(a) one embraces the interpretational approach *simpliciter*—remaining silent on a justification; or
(b) one embraces the interpretational approach alongside reduction—i.e., the interpretational approach is justified by the promised guarantee that reduction is universally *possible*, without having to actually provide a reduced theory in advance (as Dewar has remarked, there are severe problems with this—though see e.g. Caulton (2015) for a defence); or
(c) one embraces the interpretational approach alongside sophistication (as with Dewar)—i.e., the interpretational approach is justified by the promised guarantee that (external) sophistication is universally *possible*, without having to actually provide the sophisticated theory in advance; or
(d) one embraces the motivational approach alongside reduction—i.e. only once a reduced theory is provided can the symmetry-to-(un)reality inference proceed (since Dasgupta rejects sophisticated substantivalism in Dasgupta (2011), yet appears to embrace motivationalism in Dasgupta (2016), arguably this author falls into this category—see Read and Møller-Nielsen (2020a, §5.3) for further discussion); or
(e) one embraces the motivational approach alongside sophistication—i.e., only once a sophisticated version of the theory is provided can the symmetry-to-(un)reality inference proceed; or
(f) one embraces some more complicated combination of these views.

As we have seen, Møller-Nielsen in fact falls into category (f), for he invariably endorses the motivational approach to symmetries, yet thinks that reduction is appropriate only in some cases (namely, when the symmetry-related models under consideration are not isomorphic), whereas otherwise one should embrace (traditional) sophistication. We stand with Møller-Nielsen—the only differences being that (i) we do allow super-empirical criteria to block one's motivation to construct a reduced/sophisticated version of one's original theory (cf. Sect. 2.3), and (ii) we do allow that internal sophistication may afford a means of explicating the common ontology of symmetry-related models.

4 On sophistication

With all of the above in hand, we turn now to the main event: a critical evaluation of Dewar's (external) sophistication about symmetries. In the following subsections, we present what we take to be some problems with this approach to articulating the ontology of symmetry-related models of a given theory. In Sect. 4.1, we question the scope of the examples which Dewar presents in Dewar (2019) in favour of sophistication—and argue that these do *not* provide compelling motivation for universal external sophistication (and a fortiori not for the interpretational approach to symmetries). In Sect. 4.2, we question the extent to which sophistication can preserve explanatory power—we argue that this is not *invariably* the case. The discussions in

these first two subsections allow us to develop in more detail some general concerns regarding external sophistication—concerns which we express in Sect. 4.3.

4.1 On universality

In Dewar (2019, §4), several examples of sophistication are presented, before it is proposed that sophistication—specifically external sophistication—can be applied *universally* to articulate the ontology of symmetry-related models. If these examples were representative, this would go some way towards rendering plausible the generalisation of the applicability of external sophistication to all cases of symmetries, as well as illustrating in more detail what is meant by this form of sophistication. It is, however, unclear whether the examples are truly successful in this regard.

Dewar's simplest example is *instantaneous electrostatics in terms of potentials* (2019, pp. 489, 494). In this case (for reasons which will become clear), it will serve to be more explicit than hitherto about the structure of the KPMs of this theory. These (in order to follow Dewar's understanding of the theory) we take to be sextuples $\langle M \times \mathbb{R}, \delta_{ij}, \epsilon_{\mathsf{ab}}, \gamma, \phi^{\mathsf{a}}, \rho \rangle$, where δ_{ij} is a Euclidean metric on the three-dimensional differentiable manifold M, and ϕ and ρ represent, respectively, the electrostatic potential and charge density. In addition, we have included the 'internal' manifold $\mathbb{R}$ in which ϕ takes its values, the Euclidean metric ϵ_{ab} on this $\mathbb{R}$, an 'internal' scalar field γ which picks out the origin of this space, and 'internal' indices on ϕ (used to specify the value of ϕ in $\mathbb{R}$ at a particular $p \in M$). DPMs of this theory are picked out via the dynamical equation

$$\delta_{ij} \nabla^i \nabla^j \phi = 4\pi\rho. \tag{5}$$

The semantics of this theory are such that ϕ takes value in $\mathbb{R}$. Then, adding any constant κ to any ϕ that solves this equation generates a new solution. If we assume that all solutions differing merely by such a κ-shift are empirically equivalent, then this κ-shift constitutes a symmetry transformation (of the kind relevant to this paper—cf. Sect. 2.2). Do these symmetry-related models represent the same *physical* state of affairs? This question is analogous to the shift arguments in NGT set on Newtonian spacetime—albeit with an internal, rather than external (i.e., spacetime) transformation.[27]

Consider a map $\psi : \phi^{\mathsf{a}} \mapsto \phi^{\mathsf{a}} + \kappa^{\mathsf{a}}$ on the internal space which implements a constant shift of the value of ϕ^{a}. Such a map can be used to generate a new model of this theory, $\mathcal{M}_{\text{shift}} = \langle M \times \mathbb{R}, \delta_{ij}, \epsilon_{\mathsf{ab}}, \gamma, \psi_* \phi^{\mathsf{a}}, \rho \rangle$. Since the diffeomorphism acts only on the internal space, we have $\delta_{ij} = \psi_* \delta_{ij}$ and $\psi_* \rho = \rho$; moreover, since these transformations are a symmetry of the Euclidean metric ϵ_{ab} on $\mathbb{R}$, we have $\epsilon_{\mathsf{ab}} = \psi_* \epsilon_{\mathsf{ab}}$. However, since these transformations shift the origin on $\mathbb{R}$, as given by

[27] There is a sense in which the case is analogous to the static shift: to ϕ is added a constant factor κ. But there is also a sense in which the case is analogous to the kinematic shift: with γ included in the semantics of this theory, the pre- and post-shift models are *not* isomorphic. In fact, the most directly analogous Newtonian shift scenario is of static shifts in *Aristotelian spacetime* (which is Newtonian spacetime supplemented with a preferred point—see e.g. Weatherall (2018, §4) for relevant discussion), in which, for a static shift, the pre- and post-shift models are again *not* isomorphic.

γ, we have $\gamma \neq \psi_* \gamma$—meaning that such shifts of the electrostatic potential ϕ do *not* generate isomorphic models. In this case, as in the kinematic shift, a mere 'traditional sophisticationist' metaphysical move (in this case anti-quidditism—see below) is not sufficient to regard these models as representing the same physical state of affairs; rather, one must also mathematically reformulate, to excise the origin γ of $\mathbb{R}$, and move to a formulation in which ϕ^{a} is valued in a one-dimensional metric affine space. Only then does traditional sophistication suffice as a means of regarding the models as representing the same physical state of affairs.

In the case of substantivalism, this renouncing of a (primitive) non-qualitative, transworld identity of manifold points (here, points in $\mathbb{R}$) means renouncing haecceitism; in the case of electrostatics the manoeuvre goes under the name of *anti-quidditism*. The idea is that just as haecceities allow for the primitive identification of *objects*, such as spacetime points, across possibilities, so too do quiddities allow for the primitive identification of *property-holdership* across possibilities. It is important to note, though, that while (anti-)quidditism often (implicitly) refers to *determinable* properties, such as 'having mass' or 'being charged', we are here concerned with transworld identification of *determinate* properties, such as 'having *that* mass' or 'having a mass of 1kg'.[28] What is at stake in our discussion of electrostatics is not determining whether the counterpart of the electrostatic potential ϕ in a κ-shifted world is still an electrostatic potential or instead, say, a gravitational potential. The issue is whether there is a determinate magnitude of the electrostatic potential ϕ in one world—say the magnitude that is represented by ϕ having the numerical value zero—that can be identified with a determinate magnitude of the electrostatic potential ϕ in another world. In other wor(l)ds, is there a matter of fact about determinate magnitudes of the electrostatic potential, or only about differences in magnitudes? To sum up: anti-quidditism (about determinate properties) denies the existence of quidditistic facts about the (determinate) properties in our theories. Given this thesis, no meaningful sense can be made of κ-shifted ϕ as fields representing distinct possibilities. Terminology aside, however, both the anti-haecceitist and anti-quidditist strategies are clearly forms of traditional sophistication—for they identify distinct but *isomorphic* models of our physical theories.

So: on Dewar's understanding of the electrostatics case, κ-shifted models are *not* isomorphic, but one can (in our reconstruction) implement the internal sophisticationist strategy by (a) modifying the semantics of the theory in order to excise γ from its models, and (b) interpreting the points in $\mathbb{R}$ anti-quidditistically. It is worth noting here, however, that there is another understanding of electrostatics—different from Dewar's—in which the origin γ is not included in the semantics, and transworld comparison of (field values at) points in $\mathbb{R}$ is facilitated entirely by a quidditistic understanding of the points in that manifold; in that case (just as in the static shift in NGT), no mathematical reformulation is necessary in order to understand these models as representing the same physical state of affairs; rather, an anti-quidditist metaphysical move suffices. (Indeed, there is a sense in which including the origin γ of $\mathbb{R}$ is unnatural, as one must then make *two* conceptually similar philosophical

[28] Or, more correctly, 'having the mass that is in our world represented by the numerical quantity 1kg'. One should be careful not to misinterpret (Dewar 2019, pp. 504–505) as suggesting that it is determinable rather than determinate properties which are relevant here.

moves in order to regard models of electrostatics differing at most by κ-shifted ϕ as being physically equivalent; for this reason, a version of the theory *without* inclusion of γ is our preferred initial formulation of the theory.) In either case, though, it should be clear that what Dewar has illustrated here is an instance of *internal* sophistication (whether involving (a) both semantic reformulation and traditional sophistication, as on the former version of the theory, or (b) just traditional sophistication, as on the latter); the example provides no positive illustration of how external sophistication is supposed to work. Indeed, this is especially so as this particular example merely excises structure naturally associated with quidditistic differences: it remains unclear, absent further details, how the sophisticationist strategy is supposed to work in more complex cases of symmetries.

One can understand Dewar's second example—of full-blown electromagnetism, in its vector potential formulation—as being intended to address this concern. KPMs of this theory (eschewing an explicit presentation of the 'internal' machinery, which we included in the previous example) are quadruples $\langle M, \eta_{ab}, A^a, J^a \rangle$, where η_{ab} is a fixed Minkowski metric field on the four-dimensional differentiable manifold M, A^a is a four-vector encoding Maxwell fields, and J^a is a source term. DPMs of this theory are the Maxwell equations,

$$\nabla_a \left(\nabla^a A^b - \nabla^b A^a \right) = J^b, \tag{6}$$

where ∇ is the derivative operator compatible with η_{ab}. It is very well-known that (6) is invariant under electromagnetic 'gauge transformations' of the form

$$A^a \mapsto A^a - \nabla^a \Lambda, \tag{7}$$

for some scalar function Λ; moreover, models of electromagnetism related by such gauge transformations are typically taken to be empirically equivalent. In this case (as in Dewar's understanding of electrostatics), models related by (7) are *not* isomorphic: the transformation under consideration is the analogue of the Leibnizian kinematic shift, rather than the static shift. So: how does Dewar's sophisticationist strategy proceed in the context of this theory?

As before, in this case, Dewar does more than merely declare that these symmetry-related models may be treated as being isomorphic, as per the external sophisticationist agenda. In addition, Dewar *mathematically reformulates* the theory, such that the correlates of the symmetry-related models in the original version of electromagnetism are indeed isomorphic in the reformulated theory. In this case, Dewar proposes that recourse to the machinery of fibre bundles is the appropriate strategy:

> Finally, consider the electromagnetic theory. This time, models of the theory are to be connections on a principal $U(1)$-bundle over $\mathbb{R}^4$. [Footnote suppressed.] Once more, we retain [(6)], but now interpreted in a way that makes use only of the more minimalist structure available in the models: $[A^a]$ is now interpreted as the vector potential of the target connection relative to some arbitrarily chosen flat connection on the principal bundle. (Dewar 2019, p. 501)

Note that here, Dewar is explicitly presenting a different (more impoverished) mathematical formalism for electromagnetism—that is, a *different* space of kinematic (a fortiori dynamic) possibilities. In this new formalism, solutions of the theory related by (7) *are* isomorphic—so that one may (finally—as in the previous example) apply traditional sophistication in order to regard all such models as representing the same physical state of affairs.[29]

We have absolutely no qualms about this overall (internal sophisticationist) strategy (beyond issues raised in Sect. 4.2) as affording, at least prima facie, a novel tool for articulating the ontology of symmetry-related models, compared to reduction. However, it is important to note that the mathematical reformulation here—just as in the case of reduction when faced with non-isomorphic models—was *essential* to the project: external sophistication of the theory under consideration was *insufficient* to do this, absent carrying through the project of also mathematically reformulating the theory under consideration—that is, absent the project (just as in the case of reduction) of finding a new formalism via which models of the original theory may ultimately be interpreted. Given that, in general, it may be very difficult to identify what the appropriate mathematical reformulation is supposed to consist in (a fact which Dewar also acknowledges: "it is often very opaque what kind of internal construction will correspond to an external construction." Dewar (2015, p. 503)), in our view, the external sophisticationist strategy is *insufficient* as a means of articulating the ontology of symmetry-related models. Moreover, this difficulty highlights that in this case there was no guarantee that such a formulation was even possible, and a fortiori there is no universal generalisation of such a guarantee for all symmetries in all theories. Sometimes one first needs to explicitly perform non-trivial mathematical feats, such as moving to the fibre bundle formalism of electromagnetism. Without such a universal guarantee, this otherwise-promising strategy of internal sophistication is most naturally paired with motivationalism (i.e. options (e) or (f) in Sect. 3.2), instead of providing a justification of interpretationalism (option (c).[30]

Finally, consider the third example presented by Dewar, which is posed in the language of first-order predicate logic. 'Kinematically', the objects of this theory are two predicates, L and R; 'dynamically', the following two sentences are satisfied:

$$\forall x(Lx \vee Rx), \tag{8}$$

$$\forall x \neg (Lx \wedge Rx). \tag{9}$$

It is true that if we have a solution to this theory, swapping all L and R predicates—that is, predicating (only) L of all objects originally instantiating R and vice versa—takes a solution to another solution, i.e. preserves the 'dynamics'. What is not the case is that this suffices to call such a swap a 'symmetry' in the relevant sense (cf. Sect. 2.2), despite Dewar's claim to the contrary (Dewar 2019, p. 488). Without any further context, the predicates L and R could be anything—and so it need not be the case that the models related by the L/R-switch are empirically equivalent, which, as we have

[29] See Weatherall (2016) for a clarification that, in this case, neither reduction nor internal sophistication yields a theory with surplus structure.

[30] We are grateful to Caspar Jacobs for discussion on the contents of this paragraph.

seen, is the sense of 'symmetry transformation' *relevant* to these considerations. (To take an example, let R stand for 'being a rhinoceros' and L stand for 'being a leopard'. A world with one rhino and six leopards is not empirically equivalent to a world with one leopard and six rhinos!) In any case, let us restrict to the empirically equivalent L/R-swaps, as seems to be what Dewar has in mind when he makes the suggestion—which he presents as a mere heuristic whilst it thus actually does a lot of work—that we think of L and R as referring to the 'left-handedness' and 'right-handedness' of gloves (Dewar 2019, p. 487). (Here we should presumably further assume parity conservation, as would be true of gloves governed by, say, NGT.) As in the electrostatics case, Dewar proposes that sophistication in this case can proceed via the rejection of quiddities (Dewar 2019, pp. 498–499).[31] According to such anti-quidditism about determinate handedness, there is no way of identifying the handedness of gloves across worlds; they are after all (at least intrinsically) qualitatively identical (*pace* Van Cleve 1987[32]), since the shape and size of all the components and the angles between them are identical between gloves. In other words, there is no absolute, primitive sense in which a glove is 'left-handed' beyond it being enantiomorphically related to what we conventionally call, in that same world, a 'right-handed' glove (i.e., there is no primitive, non-qualitative, non-conventional determinate property of 'being left-handed' to be instantiated beyond such a pair of incongruent gloves each instantiating the determinable, qualitative property 'being handed' as well as the determinate, qualitative property 'standing in the opposite configuration as the other glove'.) If this is what Dewar proposes, though, then this analysis is clearly once again (at best[33]) an instance of *traditional* sophistication—non-qualitative distinctions between isomorphic worlds are simply 'forgotten'.

If Dewar's examples were illustrative, representative instances of external sophistication, then this would go some way towards supporting the proposal that symmetry-related models may invariably be externally sophisticated, thereby in turn supporting the interpretational approach. Instead, however, we have seen that Dewar's examples

[31] Is quidditism about determinate handedness conceivable and metaphysically possible in the first place? If not, there would be no reason to praise Dewar's proposal for revealing such quiddities to be redundant and for getting rid of them. Perhaps Earman's primitive internal relations 'standing in a left-configuration' and 'standing in a right-configuration' could be interpreted as providing such quiddities Earman (1991) (see also Van Cleve 1987), and/or his primitive intrinsic properties R^* and L^* (Earman 1989, §7.3). Another example would be Walker's primitive orientations (Walker 1987). However, in following this tradition, Dewar ignores that it has been realised long ago (*pace* (Earman 1991, pp. 133–134)) that 'left-handedness' is not and could not be a primitive, non-qualitative, non-conventional property which allows us to compare handedness across worlds (Frederick 1991; Gardner 1990; Jammer 1960; Nerlich 1994; Van Cleve and Frederick 1991). One way of seeing this is that, if space is non-orientable, a left-handed glove and a right-handed glove could be made congruent. Even if space is orientable, a rotation in the fourth dimension would equally turn a supposedly left-handed glove into a right-handed one. Primitive non-qualitative properties should not depend on mere rigid transportations within a single possible world of the objects instantiating those quiddities. Baptising one glove as 'left-handed'—and thereby also all others that are congruent with it—is a mere linguistic convention that has no metaphysical bite, especially not in other worlds than the one in which the baptism occurred. (Moreover, if, *per impossible*, these supposed quiddities were part of the essence of being left-handed, it would have been hard to see how we could have decided to 'just' forget about them.) In other words, gloves are *born* (traditionally) sophisticated, despite the misleading notation of (8) and (9). No notion of sophistication at all plays a role in this example, let alone external sophistication.

[32] See also Huggett (2000), Martens (2011).

[33] See footnote 31.

provide no motivation at all for external sophistication: either they are merely cases of traditional sophistication, or they explicitly appeal to mathematical reformulation (and so internal sophistication), which is a strategy more naturally paired with motivationalism than with interpretationalism. Of course, none of this proves that there could be *no* illustrative examples of external sophistication, nor does it provide any argument *against* external sophistication. Discussion of these more directly critical matters is deferred to Sect. 4.3.

4.2 On explanation

As discussed in Sect. 3.1, Dewar acknowledges the importance of retaining explanatory power in the context of symmetry-to-(un)reality inferences. In fact, we are in agreement that it is a problematic feature of some reduced theories that they exhibit explanatory deficits relative to the original theories from which they are derived (Dewar 2019; Martens 2018b). Consider, for instance, the reduced theory corresponding to the above example of electrostatics, where the invariant mathematical structure is the electric field $E_i := \nabla_i \phi$. In order to capture the full content of the original theory, it is not sufficient to substitute this definition of the electric field into (5), to obtain

$$\delta_{ij} \nabla^i E^j = 4\pi\rho; \tag{10}$$

rather, one also needs to add the condition

$$\epsilon_{ijk} \nabla^j E^k = 0. \tag{11}$$

In the original theory, however, the equivalent expression is a mathematical identity:

$$\epsilon_{ijk} \nabla^j \nabla^k \phi = 0. \tag{12}$$

Thus, in our original formulation of electrostatics, equation (12) is an "analytic or definitional necessity rather than a "mere" law" (Dewar 2019, p. 497)—as it is in the case of equation (11) in our 'reduced' formulation of electrostatics. Dewar correctly points out that such a definition counts as a proper explanation on many popular accounts of explanation in philosophy of science; an explanation that is lacking in the reduced theory. The question to be considered now, then, is whether (externally) sophisticated theories *invariably* preserve explanatory virtues as compared with the unsophisticated theories from which they are constructed.

Dewar's motivation for seemingly promising as much stems from the fact that sophistication does not change the equations of a given theory—it modifies only the semantics, while leaving the syntax untouched. In the case study of electrostatics from Sect. 4.1, for example, sophistication merely replaces $\mathbb{R}$ by a one-dimensional, oriented, metric affine space as the range of ϕ. This leaves $\nabla_i \phi$ well-defined and invariant, such that equations (5) and (12) also still hold and remain well-defined, thereby preserving explanatory power. The natural question to ask at this point is, then, the following: can explanatory power arise only from the dynamical equations

of a given theory? In this subsection, we give a negative answer to this question. Explanatory power may be preserved in examples such as electrostatics—but (we contend) such examples are again unrepresentative.

Our first problem case for the claim that sophisticated theories invariably preserve explanatory power can be illustrated by considering NGT set on Newtonian spacetime. Within this theory, absolute velocities are not "idly turning wheels" (Møller-Nielsen 2017, p. 1263), but are used to *define* the observable relative velocities. Following Dewar's earlier claim about definitions, this should count as a proper explanation: absolute velocities 'indirectly' explain observable phenomena by defining/explaining relative velocities which are observable. It is then somewhat surprising that Dewar does not promise or show that the explanatory power will not decrease when such symmetry-related models are (externally) sophisticated, but instead denies that these putative explanations are truly indispensable (Dewar 2015, p. 322).

A second problem case for the claim that sophisticated theories invariably preserve explanatory power is also drawn from Newtonian gravitation. Consider force-based Newtonian gravitation, in terms of the standard set of initial variables and parameters, i.e. distance r, velocity v, and mass m. For simplicity, focus on models with two equally massive particles with zero total angular momentum. One aspect of the behaviour of these particles is encapsulated in the escape velocity formula

$$v_e = \sqrt{\frac{2Gm}{r}}. \tag{13}$$

If the initial relative velocity $v_0 := v\,(t = 0)$ of the particles is larger than v_e, they will end up escaping each other; otherwise, they will end up colliding. Two sets of initial conditions agreeing on their mass ratios can nevertheless be distinct if these ratios hold in virtue of distinct quidditistic absolute masses. Can these absolute masses make a difference? They can. The following transformation will, for an appropriate value of the scalar α, change whether the escape velocity inequality is satisfied:

$$\begin{aligned} m &\mapsto \alpha m, \\ r &\mapsto r, \\ v &\mapsto v. \end{aligned} \tag{14}$$

These non-qualitative mappings of initial states lead to empirically distinguishable evolutions—escape versus non-escape. (This transformation is thus not a symmetry (Martens 2017, 2019a, b)—*pace* Dasgupta 2013). Although this does not make absolute masses detectable in the same sense as relative velocities, it does make them detectable in some weaker sense, that still makes them more empirically relevant than, say, absolute velocities (Martens 2019b). Quidditistic absolute masses may thus explain detectable phenomena.

One may retort that since transformation (14) is not a symmetry, it has no relevance to the current project concerning symmetry-to-(un)reality inferences. In response to this concern, we proceed now by considering another theory for which transformation (14) *is* a symmetry, but in which absolute masses nevertheless play an indispensable explanatory role.

Consider the following Newtonian theory (see Martens 2019a for further discussion of this theory). (Although this theory is of course related to standard Newtonian gravity—both theories are in fact empirically equivalent—it is syntactically distinct from standard Newtonian gravity; here we will consider this theory on its own.) The gravitational force between two masses is given by

$$F_{\text{grav}} = \gamma \frac{Mm}{r^2 \sum_k m_k} = \gamma \frac{M}{r^2 \sum_k \frac{m_k}{m}}, \tag{15}$$

with γ a constant such that in the actual world $\gamma = G \sum_k m_k$. The corresponding escape velocity depends only on mass ratios—

$$v_e = \sqrt{\frac{\gamma}{r \sum_k \frac{m_k}{m}}}, \tag{16}$$

—thereby rendering transformation (14) a symmetry (in the relevant sense—cf. Sect. 2.2) of this theory; uniform mass scalings do not lead to an empirical difference. Note, moreover, that this symmetry relates qualitatively indistinguishable models.

There is some ambiguity as to how one should proceed with sophisticating this theory, as there are at least two options available for 'forgetting structure' (without modifying the syntax).[34] The most obvious option is to throw away absolute masses altogether. It was for this exact reason—obtaining an ontology with no absolute masses but only mass ratios m_{ij} (i.e., primitive and symmetric relations of 'comparative masshood' holding between particles i and j)—that this theory was designed in the first place (Martens 2019a). However, if the fundamental ontology in this theory consists only of mass relations, there is nothing to ensure the transitivity of those mass relations,

$$m_{ij} = m_{ik} \cdot m_{kj}, \quad \text{for any three particles } i, j, k. \tag{17}$$

To make matters worse, without such a transitivity constraint holding, one could not even coherently interpret these mass relations as ratios in the first place (Martens 2019a). Only if one commits to absolute masses in virtue of which the mass ratios hold does one explain the transitivity of those mass relations, as the following becomes a mathematical truth (Armstrong 1988; Martens 2018a, 2019a; Roberts 2016; Russell 1903):

$$\frac{m_i}{m_j} = \frac{m_i}{m_k} \cdot \frac{m_k}{m_j} \quad \text{for any three particles } i, j, k. \tag{18}$$

Recall that Dewar agrees that this constitutes an explanation. Thus, in this theory, absolute masses may not be invariant under the symmetry transformation (14), but

[34] Note that we think of absolute masses and mass relations/ratios in the sense defined by Martens (2019b, §3), not in terms of property spaces.

they are nevertheless explanatorily indispensable: they ensure transitivity of mass ratios as a matter of mathematical fact. It is hard to see how any other ontology could dispense with absolute masses whilst retaining this explanation.

This pushes one towards the second option for sophisticating this theory: retain the absolute masses but 'forget' just their quiddities. (As with the previous option, the syntax is not modified.) As Jacobs argues, this suffices to explain the transitivity of mass relations (Jacobs 2019). This therefore seems the best option for a weak interpretationalist. It should be noted though, once more, that the original motivation for introducing (15) was to dispose of absolute masses in virtue of which the mass ratios obtain.[35] It goes against the spirit of interpretationalism to retain more than what is invariant under the symmetry. Moreover, this confirms once again that explanatory considerations—mass ratios obtain in virtue of, i.e. are metaphysically explained by, underlying absolute masses (albeit non-quidditistic ones)—that go beyond syntax and dynamical equations can play a crucial role in determining the metaphysical consequences of symmetries. Furthermore, even if this option is indeed the best of both options, the fact remains that a non-trivial choice between two options had to be made. The whole point of external sophistication and of interpretationalism is that one is supposed to be able to draw immediate metaphysical consequences from the symmetries of a theory. The fact that we needed to do some work (cf. our discussion of gauge transformations within electromagnetism in Sect. 4.1), that we needed to make non-trivial decisions before the appropriate ontology was revealed, resonates more with motivationalism than interpretationalism.

Let us finally and briefly return to another problematic symmetry: that of gauge shifts of the electromagnetic vector potential (Sect. 4.1), but specifically in the context of the Aharonov-Bohm effect (Aharonov and Bohm 1959). This context has been discussed frequently and in detail in the literature—e.g. Dewar (2019), Healey (2007), Martens (2018b)—so we will not dwell on it here. Suffice it to say that, although we agree that the sophisticated theory can recover analogues of (12) and (17), to the extent that this counts as an explanation it is not an explanation that is, in all relevant senses, separable (*pace* Jacobs 2019), as Dewar acknowledges (Dewar 2019, fn. 56). At the same time, there is a reduced theory available that is local in all relevant senses (Wallace 2014).[36]

To sum up: the promise of universal external sophistication that explanatory powers will invariably be preserved is motivated by the fact that such sophistication leaves the dynamical equations untouched. We agree that this ensures that explanations provided by those dynamical equations are preserved, and we have seen that this constitutes an improvement upon reductionism for some cases. However, this motivation remains silent on other types of explanation (as well as on other super-empirical criteria). Dewar has thus not provided a *universal* argument for explanatory power being preserved after sophistication. The examples in this subsection indicate that there cannot be such argument, as sophistication fails to invariably preserve explanations that do not derive from the dynamical equations. On both the weak interpretational approach and

[35] This view goes under the name of comparativism, i.e. the denial of absolutism, about mass (Dasgupta 2013; Martens 2019b).

[36] For further discussion of these issues in the context of the debates which are the focus of this paper, see Martens (2018b).

the strong motivational approach to symmetry-to-(un)reality inferences, loss of such explanatory power may then be considered as a reason for blocking these inferences, even when these approaches are (partially) combined with (traditional or external) sophistication.

4.3 On sophistry

If the examples presented by Dewar (2019), discussed in Sect. 4.1, do not support his proposal for the universal application of external sophistication, might there be independent positive reasons for believing in this proposal, as there are for the restricted application of sophistication (i.e., the traditional notion)? Or (weaker still), can any transparent examples of external sophistication of non-isomorphic symmetry-related models be given?

We are sceptical. To *stipulate* that qualitatively distinct, i.e. non-isomorphic models, such as models of NGT related by kinematic shifts, are nevertheless isomorphic reads prima facie as nothing more than a flat-out contradiction.[37] The burden falls on the advocate of external sophistication to explicate what (s)he has in mind here—for to simply *insist* that all symmetry-related models be regarded as being isomorphic simply appears to be begging the interpretative question—of Russellian theft over honest toil (Russell 1919, p. 71).

We submit that Dewar is seeking to have his cake and eat it. He cannot. One natural worry to have about the external sophisticationist programme is that it is not (fully) realist in spirit to begin with.[38] As is evident from Sect. 3.1, Dewar suggests that one can be a realist (simpliciter) without having to make any commitments as to which parts of a theory one is realist or anti-realist *about*—that is, without having to commit to realism about anything *specific*. This is certainly consonant with external sophistication—but in that case, it is often opaque what reality is being subscribed to; moreover, it is often unclear what grounds or explains or justifies the physical equivalence of models that, on a natural interpretation, represent distinct possible worlds (Møller-Nielsen 2017). Dewar concedes as much when he states that it is "often very opaque" (Dewar 2019, p. 503) what kind of semantics (if any!) corresponds to this stipulation that symmetry-related models should be considered 'as if' they are isomorphic. We confess, in line with Møller-Nielsen (2017) and Sider (2018, §5), that we find it difficult to make sense of such an opaque realism.

[37] Perhaps Dewar has in mind a decoupling of the notions of 'qualitatively identical' and 'isomorphic', with kinematic shifts producing models that are not qualitatively identical but nevertheless isomorphic. But what could this concept of isomorphism be? 'Identical in structure up to empirically indistuinguishable structure'? That would amount to verificationism, a label that Dewar resists (Dewar 2015, p. 320) (and for good reason, as his proposal would thereby fail to connect to the realist project that both the interpretational and motivational approaches purport to be involved in—more on this below). As a general point: we take it that one reason to favour traditional sophistication or reduction over external sophistication is that the former two interpretative strategies are less prone to collapse into verificationism—we thank an anonymous reviewer for pushing us on this.

[38] Note that this objection against taking sophistication to be a realist-approved stratagem is distinct from other such objections towards the same conclusion which have already been responded to by Dewar (2015, §6).

In this context, it is helpful to emphasise the following belief which we do share with Dewar:

> if we want to know the answers to specific questions about the nature of a theory's ontology and ideology, then [a reformulated theory] is invaluable. (Dewar 2015, p. 326)

We take it that a complete and honest form of realism should not only take these questions—such as the question of which parts of a theory one is to be realist about and which parts one is not—on board, but consider them to be crucial. Leaving them out is at best a dishonest form of realism, and at worst a form of anti-realism. Treating qualitatively distinct models related by a symmetry 'as if' they are isomorphic does not help one do the hard interpretative work that is required of realism.

5 Conclusions

Dewar's work on the philosophy of symmetries has precipitated significant advances in the field; moreover, the concept of internal sophistication is an important new tool which can (and should) be deployed in consideration of symmetry-to-reality inferences. This notwithstanding, however, we are sceptical of the notion of *external* sophistication—which, to our minds, does not attain a sufficiently high level of metaphysical perspicuity in order to constitute a means of explicating the ontology of symmetry-related models of a given theory. In Dewar (2019), Dewar has motivated only traditional sophistication (which is *already known* to be a viable metaphysical thesis) and internal sophistication (which involves more than merely forgetting structure, as it also requires mathematical reformulation); he has not motivated external sophistication. Finally, while it is true that sophistication (insofar as the proposal makes sense) *can* preserve the explanatory merits of the original theory under consideration (whereas a reduced theory will often lack those merits), we have seen in this paper reasons to doubt that sophistication *invariably* preserves explanatory qualities. Our conclusions are, therefore, threefold: (a) as a thesis purporting to shed light on the ontology of symmetry-related models, external sophistication about symmetries is lacking; (b) one must be cautious when it comes to making claims about the explanatory merits of sophisticated theories; and (c) the proposal of universal external sophistication, as it stands, thereby a fortiori fails to prove that it can provide the support that the interpretational approach to symmetries requires.

Acknowledgements Open Access funding provided by Projekt DEAL. We thank Neil Dewar for countless discussions on symmetries, and for his patient and constructive engagement with our work. In addition, we thank Tom Møller-Nielsen for the invaluable role he has played in shaping this debate, and in paving our way. We are also very grateful to Caspar Jacobs, and to the Cambridge Simplex, for valuable feedback. Finally, N.M. thanks the DFG Research Unit "The Epistemology of the Large Hadron Collider" (Grant FOR 2063) for their support over the course of writing this paper.

References

Aharonov, Y., & Bohm, D. (1959). Significance of electromagnetic potentials in quantum theory. *Physical Review*, *115*, 485–491.

Anderson, J. L. (1967). *Principles of relativity physics*. New York and London: Academic Press.

Armstrong, D. (1988). Are quantities relations? A reply to bigelow and pargetter. *Philosophical Studies*, *54*, 305–316.

Belot, G. (2003). Symmetry and gauge freedom. *Studies in History and Philosophy of Modern Physics*, *34*, 189–225.

Belot, G. (2013). Symmetry and equivalence. In R. Batterman (Ed.), *The Oxford handbook of philosophy of physics* (pp. 318–339). Oxford: Oxford University Press.

Belot, G. (2018). Fifty million Elvis fans can't be wrong. *Noûs*, *52*(4), 946–981.

Brading, K., & Castellani, E. (Eds.). (2003). *Symmetries in physics: Philosophical reflections*. Cambridge: Cambridge University Press.

Brading, K., & Teh, N. J. (2017). Symmetry and symmetry breaking. In: E. N. Zalta (Ed.). The Stanford Encyclopedia of Philosophy.

Brown, H. R. (2005). *Physical relativity: Space-time structure from a dynamical perspective*. Oxford: Oxford University Press.

Butterfield, J. (2018). On dualities and equivalences between physical theories. Forthcoming in N. Huggett, B. Le Bihan, & C. Wüthrich (Eds.), *Philosophy beyond spacetime*. Oxford: Oxford University Press 2020.

Caulton, (2015). The role of symmetry in the interpretation of physical theories. *Studies in History and Philosophy of Modern Physics*, *52*, 153–162.

Dasgupta, S. (2011). The bare necessities. *Philosophical Perspectives*, *25*, 115–160.

Dasgupta, S. (2013). Absolutism vs. comparativism about quantity. In K. Bennett & D. W. Zimmerman (Eds.), *Oxford studies in metaphysics* (Vol. 8, pp. 105–147). Oxford: Oxford University Press.

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

De Haro, S., & Butterfield, J. (2018). A schema for dualities, illustrated by Bosonization. In J. Kouneiher, et al. (Eds.), *Foundations of mathematics and physics one century after Hilbert*. Berlin: Springer.

Dewar, N. (2015). Symmetries and the philosophy of language. *Studies in the History and Philosophy of Modern Physics*, *52*, 317–327.

Dewar, N. (2016). *Symmetries in physics, metaphysics, and logic*, D.Phil. thesis, University of Oxford.

Dewar, N. (2019). Sophistication about symmetries. *British Journal for the Philosophy of Science*, *70*, 485–521.

Dirac, P. (1930). *The principles of quantum mechanics*. Oxford: Oxford University Press.

Earman, J. (1991). Kant, incongruous counterparts, and the nature of space and space-time. In J. van Cleve & R. E. Frederick (Eds.), *The philosophy of right and left: Incongruent counterparts and the nature of space* (pp. 131–150). Dordrecht: Kluwer. (Originally published in *Ratio*, 13 (1971), pp. 1–18).

Earman, J. (1989). *World enough and space-time: Absolute versus relational theories of space and time*. Cambridge, MA: MIT Press.

Esfeld, M., & Lam, V. (2011). Moderate structural realism about spacetime. *Synthese*, *160*(1), 27–46.

Frederick, R. E. (1991). Introduction to the argument of 1768. In J. van Cleve & R. E. Frederick (Eds.), *The philosophy of right and left: Incongruent counterparts and the nature of space* (pp. 1–14). Dordrecht: Kluwer.

Gardner, M. (1990). *The new ambidextruous universe: Symmetry and asymmetry from mirror reflections to superstrings* (3 revised ed.). New York, NY: W.H. Freeman. (First edition (*The Ambidextruous Universe*) published in 1964).

Gryb, S., & Thébault, K. P. Y. (2016). Time remains. *British Journal for the Philosophy of Science*, *67*, 663–705.

Healey, R. (2007). *Gauging what's real*. Oxford: Oxford University Press.

Huggett, N. (2000). Reflections on parity nonconservation. *Philosophy of Science*, *67*, 219–241.
Ismael, J., & van Fraassen, B. (2003). Symmetry as a guide to superfluous theoretical structure. In K. Brading & E. Castellani (Eds.), *Symmetries in physics: Philosophical reflections* (pp. 371–392). Cambridge: Cambridge University Press.
Jacobs, C. (2019). *Gauge and explanation: Can gauge-dependent quantities be explanatory?* B.Phil. thesis, University of Oxford.
Jammer, M. (1960). *Concepts of space*. New York: Harper.
Malament, D. B. (2012). *Topics in the foundations of general relativity and newtonian gravitation theory*. Chicago, IL: University of Chicago Press.
Martens, N. C. M. (2011). *Parity violation and the reality of space*, B.A. Thesis, University of Groningen.
Martens, N. C. M. (2017). *Against comparativism about mass in Newtonian gravity: A case study in the metaphysics of scale*, D.Phil. thesis, Magdalen College, University of Oxford.
Martens, N. C. M. (2018a). Against Laplacian reduction of Newtonian mass to spatiotemporal quantities. *Foundations of Physics*, *48*(5), 591–609.
Martens, N. C. M. (2018b). Symmetry-to-(un)reality inferences & explanatory power: The case of the Aharonov-Bohm effect. Unpublished draft.
Martens, N. C. M. (2019a). Machian comparativism about mass. *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axz013.
Martens, N. C. M. (2019b). The (un)detectability of absolute Newtonian masses. *Synthese*. https://doi.org/10.1007/s11229-019-02229-2.
Martens, N. C. M., & Lehmkuhl, D. (2019). Dark Matter = Modified Gravity? Scrutinising the spacetime-matter distinction through the modified gravity/dark matter lens. (Unpublished draft).
Matsubara, K. (2013). Realism, underdetermination and string theory dualities. *Synthese*, *190*, 471–489.
Møller-Nielsen, T. (2015). *Symmetry, indiscernibiliy, and the generalist picture*, D.Phil. thesis, Balliol College, University of Oxford.
Møller-Nielsen, T. (2017). Invariance, interpretation, and motivation. *Philosophy of Science*, *84*(5), 1253–1264.
Nerlich, G. (1994). *The shape of space* (2nd ed.). Cambridge: Cambridge University Press.
Norton, J. (2019). The hole argument. In: E. N. Zalta (Ed.). *The Stanford encyclopedia of philosophy*.
Nozick, R. (2001). *Invariances: The structure of the objective world*. Cambridge, MA: Harvard University Press.
Polchinski, J. (2017). Dualities of fields and strings. *Studies in History and Philosophy of Modern Physics*, *59*, 6–20.
Pooley, O. (2002). *The reality of spacetime*, D.Phil. Thesis, Balliol College, University of Oxford.
Pooley, O. (2015). *The reality of spacetime*, book draft.
Read, J. (2016). The interpretation of string-theoretic dualities. *Foundations of Physics*, *46*(2), 209–235.
Read, J., & Møller-Nielsen, T. (2020a). Motivating dualities. *Synthese*, *197*, 263–291.
Read, J., & Møller-Nielsen, T. (2020b). Redundant epistemic symmetries. *Studies in History and Philosophy of Modern Physics*. https://doi.org/10.1016/j.shpsb.2020.03.002.
Rickles, D. (2008). *Symmetry, structure, and spacetime, philosophy and foundations of physics* (Vol. 3). Amsterdam: Elsevier.
Rickles, D. (2011). A philosopher looks at string dualities. *Studies in History and Philosophy of Modern Physics*, *42*, 54–67.
Roberts, J. T. (2016). *A case for comparativism about physical quantities*. https://www.academia.edu/28548115/A_Case_for_Comparativism_about_Physical_Quantities_--_SMS_2016_Geneva.
Russell, B. (1903). *The principles of mathematics*. Cambridge: Cambridge University Press.
Russell, B. (1919). *Introduction to mathematical philosophy*. New York and London: George Allen & Unwin Ltd.
Saunders, S. (2003). Physics and Leibniz's principles. In K. Brading & E. Castellani (Eds.), *Symmetries in physics: Philosophical reflections*. Cambridge: Cambridge University Press.
Sider, T. (2018). *The tools of metaphysics and the metaphysics of science*, book draft.
Van Cleve, J. (1987). Right, left, and the fourth dimension. *The Philosophical Review 96*, 33–68. Reprinted in Cleve, J. van, Frederick, R. E. (Eds.). *The Philosophy of Right and Left: Incongruent Counterparts and the Nature of Space*, Dordrecht: Kluwer Academic Publishers, pp. 203–234, 1991.
Van Cleve, J., & Frederick, R. E. (Eds.). (1991). *The philosophy of right and left: Incongruent counterparts and the nature of space*. Dordrecht: Kluwer.
Van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.

Walker, R. (1987). Incongruent counterparts. In *Kant*. London: Routledge & Kegan Paul (pp. 44–51). Reprinted in J. van Cleve, and R. E. Frederick (eds.), *The Philosophy of Right and Left: Incongruent Counterparts and the Nature of Space*, Dordrecht: Kluwer Academic Publishers, pp. 187–194, 1991.
Wallace, D. (2014). Deflating the Aharonov-Bohm Effect. arXiv:1407.5073 (Unpublished draft).
Weatherall, J. O. (2016). Understanding gauge. *Philosophy of Science*, *85*(5), 1039–1049.
Weatherall, J. O. (2018). Regarding the 'hole argument'. *British Journal for the Philosophy of Science*, *69*, 329–350.