Motivating Dualities

James Read* and Thomas Møller-Nielsen†

Word count: 12685

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

Abstract

There exists a common view that for theories related by a 'duality', dual models typically may be taken *ab initio* to represent the same physical state of affairs, i.e. to correspond to the same possible world. We question this view, by drawing a parallel with the distinction between 'interpretational' and 'motivational' approaches to symmetries.

Contents

*james.read@philosophy.ox.ac.uk
†thomas.moller-nielsen@philosophy.ox.ac.uk

1 Introduction

The phenomenon of 'duality' is pervasive in contemporary theoretical physics—particularly string theory. Roughly, two physical theories are dual when there exists an isomorphism between their spaces of dynamically possible models,[1] such that models related by that isomorphism are empirically equivalent. According to a common view in the philosophical literature, duality-related models typically may be construed *ab initio* as representing the same physical state of affairs, i.e. as corresponding to the same possible world—in which case duality-related

[1]For the definition of 'dynamically possible model', see §2. In this paper, the term 'model' is understood in the sense of the semantic conception of scientific theories—see e.g. [71, ch. 2], and §2.

models are not only empirically equivalent, but also physically equivalent. Two motivations for this view are often advanced:

(1) This view of dualities aligns with a general conception of the philosophical import of *symmetry transformations*—namely, that models related by a symmetry transformation typically may be understood *ab initio* as being physically equivalent.

(2) This view of dualities accords with a perceived consensus within the contemporary theoretical physics community.

In this paper, we question both (1) and (2). On (1), we deny that dual models may be regarded as physically equivalent absent a coherent explication of the common ontology underpinning this physical equivalence; and by the same token, we deny that symmetry-related models may be regarded as physically equivalent in the absence of such an explication. Thus, we argue for a reconstrual of the import of dualities and symmetries: dualities invariably at most *motivate one to seek* an understanding of how it is that dual models are to be regarded as physically equivalent; and by the same token, symmetries also invariably at most *motivate one to seek* an understanding how it is that symmetry-related models are to be regarded as physically equivalent. On (2), we cite a variety of evidence from the physics literature which calls into question whether this perception of such a consensus is correct.[2]

The format of this paper is as follows. In §2, we recall some of the central features of the semantic approach to scientific theories—this being the framework largely adopted in this paper. In §3, we introduce the distinction between the 'interpretational' and 'motivational' approaches to symmetries, using Newtonian gravitation theory as an illustrative example; we go on to defend the motivational approach.[3] In §4, we present notions of 'underdetermination' and 'theoretical equivalence' which will prove useful in our subsequent discussion of dualities in §5—in which we introduce a distinction between the interpretational and motivational approaches to *dualities*, and (again) defend the latter approach. Finally, in §6 we assess the extent to which the interpretational apporach—a common view in the philosophical literature—is embraced in the theoretical physics community.

[2] In this regard, we follow the methodology of [7].

[3] In this regard, this paper may be viewed as continuous with [45], offering further reasons to endorse the motivational approach, as well as providing an extended application of the interpretation/motivation distinction to the case of dualities.

2 Models and Gauge

On the semantic conception of scientific theories—introduced by Suppes [67], and famously endorsed by Van Fraassen [71, 72]—a theory is associated with a class of models.[4,5] For a given theory $\mathcal{T}$, we take the most general class of associated models to be that of 'kinematically possible models' (KPMs) $\mathcal{K}$, which consists in tuples of specified geometrical objects. For example, the KPMs of general relativity (GR) are picked out by all triples of the form $\langle M, g_{ab}, \Phi \rangle$,[6] where M is a four-dimensional differentiable manifold;[7] g_{ab} is a Lorentzian metric field on M; and Φ is a placeholder for the matter fields of the theory.

Classically, a theory $\mathcal{T}$, with KPMs $\langle M, O_1, \ldots O_n \rangle$ (where the O_i are geometrical objects), comes with a set of dynamical equations for the O_i. The KPMs of $\mathcal{T}$ in which the O_i obey those dynamical equations form a subset $\mathcal{D} \subset \mathcal{K}$, the 'dynamically possible models' (DPMs) of $\mathcal{T}$. For example, in the case of GR, only those triples $\langle M, g_{ab}, \Phi \rangle$ the geometrical objects of which satisfy the Einstein field equations[8]

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

—the dynamical equations of the theory, which relate g_{ab} to the stress-energy tensor T_{ab} of the Φ—in *addition* to the dynamical equations of the Φ, are DPMs.[9] Quantum mechanically, the story changes: some of the O_i in the KPMs of $\mathcal{T}$ are understood to be *operator-valued*; DPMs

[4]One should distinguish the claim that a given theory has an associated class of models from the (more controversial) claim that a theory should be *identified* with such a class of models. In this paper, we embrace the former, but remain agnostic on the latter.

[5]Van Fraassen identifies a model of a theory as "Any structure which satisfies the axioms of [that] theory" [71, p. 53]. In the language of this paper, it is natural to identify models in Van Fraassen's sense with dynamically possible models (see below). Following e.g. [53, 54], we understand the notion of a model in a broader sense.

[6]Throughout this paper, abstract (i.e. coordinate-independent) indices are written in Latin script, and we set $G_N = c = 1$.

[7]Two points here are in order. First, one should avoid, at this stage, asserting M to be the *spacetime* manifold, for to do so is to conflate the mathematical model under consideration with the possible world to which that model is ultimately interpreted as corresponding. Second, and relatedly, in light of the debate over the hole argument [26], it is *not* necessarily correct to interpret M as representing substantival spacetime at all—though this issue will here be set aside.

[8]These are the Einstein field equations with vanishing cosmological constant Λ. For $\Lambda \neq 0$, the field equations read $G_{ab} + \Lambda g_{ab} = 8\pi T_{ab}$.

[9]Strictly, independence of these dynamical equations from the Einstein field equations depends on the case in question—see [8, §9.3] and [44, §20.6].

are picked out as those KPMs the geometrical objects of which satisfy certain correlation functions (for relevant aspects of the structure of quantum field theory, see e.g. [24,49,66]; for further philosophical details regarding the above approach, see [58, ch. 5]).

Models of a theory $\mathcal{T}$ are interpreted as representing possible worlds. Sometimes, however, we may wish to interpret two or more distinct models as representing the *same* world. In that case, the space of KPMs $\mathcal{K}$ of $\mathcal{T}$ is partitioned into classes of 'gauge-equivalent' models—which are interpreted as representing the same world—and the multiplicity of models representing the same world is an example of a 'gauge redundancy'.[10] In the case in which the interpretation of $\mathcal{T}$ leads to gauge redundancy, we may construct a reduced space of models $\tilde{\mathcal{K}}$, in which gauge-related models are mathematically identified.[11] This in turn induces a reduced space of DPMs, $\tilde{\mathcal{D}} \subset \tilde{\mathcal{K}}$.[12]

3 Interpretation and Motivation

3.1 Two Approaches to Symmetries

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 (found—although not necessarily endorsed—in e.g. [3, 5, 11, 14, 20, 23, 29, 33, 48, 64, 80]) is the following: two models of $\mathcal{T}$ typically may be regarded *ab initio* as representing the same possible world when they are related by a *symmetry transformation*—even absent a coherent explication of their shared ontology.[13,14]

[10] It should be stressed that the term 'gauge redundancy' is deployed in this paper in a broader sense than that typically found in the physics literature, where the term is often reserved for certain 'internal' symmetries associated with Yang-Mills type theories. For philosophical discussion, see e.g. [30,74,76].

[11] For a concise expression of these points in the language of category theory, see [76,78,79].

[12] One assumes that two models cannot be gauge-equivalent if they satisfy different dynamics. While one might worry that this understanding of gauge redundancies effaces the possibility that two models with different dynamics may correspond to the same possible world (and thereby pose problems for the interpretation of *dualities*—see §4 below), this is not correct, for nothing in the above precludes the possibility that there exist other relations which may allow for *inter*-theoretic model identification.

[13] Clearly, such a claim has substance only once an appropriate definition of a 'symmetry transformation' is provided; this matter is addressed in detail below.

[14] What is meant by such an explication will be made explicit over the following subsections. This explication must cohere both internally, and with the structure of the models under consideration (it is, therefore, insufficient

According to this 'interpretational' approach [45, §2], in the presence of symmetry-related models, we are (a) typically warranted in interpreting those models as representing the same possible world—even absent a coherent explication of their common ontology; then may (but are not required to) go on to (b) identify such models, to construct a reduced space of KPMs $\tilde{\mathcal{K}}$; and finally (c) seek to explicate the ontology of the models of $\tilde{\mathcal{K}}$. This is in contrast with the 'motivational' approach [45, §2], according to which the existence of symmetry-related models first (a) motivates us to provide an explication of the shared ontology of these models; but only once such an explication is forthcoming should we (b) interpret those models as representing the same possible world; and (potentially) (c) identify those models to construct a reduced space of KPMs, $\tilde{\mathcal{K}}$.[15]

Why 'typically', in the above presentation of the interpretational view? A supporter of this view *may* impose certain further criteria for when symmetry-related models are to be regarded as physically equivalent—and so need not always *actually* interpret such models as being physically equivalent. For example, even for the interpretationalist it is plausible that not all symmetry-related models should be interpreted as corresponding to the same possible world, for consider e.g. the case of *Galileo's ship*, in which only a subsystem in a model of Newtonian mechanics is boosted—in this case, we have two symmetry-related models, which nevertheless clearly do *not* correspond to the same possible world.[16] Inserting this 'typically' clause does not obscure the interpretational view, however—for the salient point is the following: on the interpretational approach, the decision to interpret symmetry-related models as being physically equivalent *need not wait upon an explication of their shared ontology*.[17]

Clearly, if the above 'interpretational' claim, and its 'motivational' alternative, are to have substance, an appropriate definition of a symmetry transformation must be provided. Suppose first that one defines such a transformation to be one upon the O_i in the KPMs of any given theory $\mathcal{T}$, such that DPMs of $\mathcal{T}$ are always taken to DPMs. Such a definition clearly will not do, for, as Belot points out, it is much too broad: [6, p. 6]

to simply assert that the two models under consideration be interpreted as corresponding to some *arbitrary* possible world).

[15] Here, we say *potentially*, for there does not necessarily exist any pressure to construct such a $\tilde{\mathcal{K}}$. To illustrate, consider the case of models related by a hole diffeomorphism in GR: even one who interprets such models as corresponding to the same possible world is not obliged to construct such a reduced theory. Cf. §3.2.6 below.

[16] For a discussion of such issues, see [59, §2]; in this paper (modulo some brief considerations in §5.3), we set these complications aside by considering only symmetry transformations which act 'globally' upon the O_i of KPMs of $\mathcal{T}$, rather than upon proper subsystems in those models.

[17] Our thanks to Neil Dewar and an anonymous referee for helpful discussion on this point.

> Ordinarily, symmetries of theories are hard to come by. But some remarkable theories have atypically large symmetry groups. The definition above effaces this sort of distinction between theories. For if we allow arbitrary permutations of the solutions of a theory to count as symmetries, then the size of a theory's group of symmetries depends only on the size of its space of solutions.

Given this, a more nuanced definition of a symmetry transformation is required; following e.g. [11,13,33], in this paper we take this to be one of *empirical equivalence*. Accordingly, we define a symmetry transformation as follows: a symmetry of a theory $\mathcal{T}$ is any automorphism of the space of DPMs of $\mathcal{T}$, *such that models related by that transformation are empirically equivalent*.[18] By 'empirical equivalence', we in turn mean that all the structures in the models under consideration corresponding to 'physically observable data' are identical between those models—that is, that the 'empirical substructures' of these models in the sense of van Fraassen [71, p. 64] coincide.[19]

It is important to be clear that we are not *endorsing* the above epistemic definition of a symmetry transformation (or the parallel epistemic definition of a duality presented in §4.1). Rather, we are merely taking it as given in this paper that symmetry-related models are empirically equivalent, while bracketing questions such as (i) whether that criterion should constitute part of the 'correct' definition of a symmetry transformation; and (ii) whether it is universally true that symmetry transformations relate (all and) only empirically equivalent models. It is, however, worth noting that for models to even potentially be physically equivalent, they must at the very least be empirically equivalent. Thus, even if one rejects the above definition of 'symmetry', one should recognise that the 'symmetries' *relevant* to our discussion here will satisfy the condition of being empirically equivalent. For a further critical discussion of these and related issues, see [46].

On the first definition of a symmetry transformation above—*viz*., that considered and dismissed by Belot—not only would it be incorrect to interpret *ab initio* all symmetry-related models as being physically equivalent (for then all models of the theory in question would

[18] This definition of a symmetry transformation has the merit of being broadly analogous with our construal of dualities, presented in §4.1.

[19] Each of [11, 13, 33] offer more nuanced ways of cashing out the 'empirical equivalence' criterion in the above definition of a symmetry transformation—for example, Dasgupta appeals both to a notion of 'how things look' [13, §6.3], and to Quinean 'observation sentences' [13, §6.3] (for details of such observation sentences, see [56,57]).

be afforded the same interpretation), but, moreover, one would clearly *not even be motivated* to find an interpretation according to which such models are physically equivalent. Thus, the motivational approach is incompatible with such a definition of a symmetry transformation. *Prima facie*, neither of these points holds for our revised definition of a symmetry transformation, featuring the additional criterion of empirical equivalence. The reason for this is that such a definition is more restrictive—so it *might* be the case that one can argue that all symmetry-related models may be regarded *ab initio* as being physically equivalent, in line with the interpretational approach; moreover, one *is* apparently motivated to find a coherent interpretation according to which such models are physically equivalent, essentially on the grounds of Occam's razor: since any structure leading to such models being interpreted as physically distinct would not be part of the empirical substructures of those models (which are identical), that structure is variant yet undetectable—so we have good *prima facie* grounds for seeking to excise it.

3.2 Newtonian Gravitation Theory

In order to clarify and develop further the distinction between the interpretational and motivational approaches to symmetries, we consider in this section the case of Newtonian gravitation theory (NGT).[20] In §§3.2.1, 3.2.2 and 3.2.3, we introduce (respectively) the KPMs, fundamental interpretational postulates, and DPMs of NGT. In §3.2.4, we introduce three important classes of symmetries of NGT, before in §§3.2.5 and 3.2.6 discussing the interpretational and motivational approaches in the context of these classes of symmetries.

Those readers uninterested in the technical details of NGT are advised to skip straight to §3.2.4. In our view, it is necessary to spell out the technical details of this theory, because we seek to provide a fully worked out example of what it *means* to fully explicate symmetry-related models' underlying ontology. More specifically, we feel the best way of conveying what a 'full explication' (or 'transparent understanding') of the reality underlying symmetry-related models amounts to is *by analogy*. Hence, we feel, the relevant technical details of this example should be spelled out in full, even though the basic ideas can plausibly be understood without them.

[20] For rigorous presentations of this theory, see e.g. [25, 28, 36, 55].

3.2.1 Kinematically Possible Models

In its field-theoretic formulation, KPMs of NGT (set in Newtonian spacetime—see [25, pp. 33ff.], and discussion below) are picked out by tuples $\langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, \varphi, \rho\rangle$, where M is a four-dimensional differentiable manifold; t_{ab} is a temporal 'metric' field on M of signature $(1, 0, 0, 0)$; h^{ab} is a spatial 'metric' field on M of signature $(0, 1, 1, 1)$;[21] ∇_a is a derivative operator on M; σ^a is a vector field; and φ and ρ are scalar fields that represent the gravitational potential field and matter density, respectively. At the level of KPMs, the following four conditions hold:

$$h^{ab}t_{ab} = 0, \tag{3.1}$$

$$\nabla_a t_{bc} = 0, \tag{3.2}$$

$$\nabla_a h^{bc} = 0, \tag{3.3}$$

$$t_{ab}\sigma^b \neq 0. \tag{3.4}$$

We refer to (3.1) as an 'orthogonality' condition, and (3.2) and (3.3) as 'compatibility' conditions. (3.4) ensures that σ^a has a component in the temporal direction,[22] and so that the images of its integral curves may be used to represent the persisting points of absolute space.

3.2.2 Interpretative Principles

Following Malament [36, p. 252], we now introduce the following interpretive principles in NGT. Let I be an open interval in $\mathbb{R}$. Then, for all smooth curves $\gamma : I \to M$:

- γ is timelike[23] if its image $\gamma[I]$ could be the worldline of a point particle.

[21] Strictly, neither t_{ab} nor h^{ab} is a metric field—see e.g. [36, p. 250]. Insofar as they are not metric fields, t_{ab} and h^{ab} are still tensor fields of rank $(0, 2)$ and $(2, 0)$, respectively.

[22] I.e. is timelike, in the sense of footnote 23.

[23] Given any vector θ^a at a point $p \in M$, we can take its 'temporal length' to be $\left(t_{ab}\theta^a\theta^b\right)^{1/2}$. We further classify θ^a as either 'timelike' or 'spacelike', depending on whether its temporal length is positive or zero, respectively. We understand a smooth curve to be 'timelike' (respectively 'spacelike') if its tangent vectors are of this character at every point along the curve. Note that (3.4) ensures that σ^a is a timelike vector field.

- γ can be reparameterised so as to be a timelike geodesic (with respect to ∇_a) iff $\gamma[I]$ could be the worldline of a free point particle.
- Clocks record the t_{ab}-length of their worldlines.

If a particle has the image of a timelike curve as its worldline, then we call the tangent field ξ^a of that curve the 'four-velocity' field of the particle, and call $\xi^b \nabla_b \xi^a$ its 'four-acceleration' field (note that strictly this is a spacelike quantity, representing the instantaneous rate of change of the three-velocity of the body in question, as determined by an inertial observer). If the particle has a mass m, then its four-acceleration field satisfies

$$F^a = m\xi^b \nabla_b \xi^a, \tag{3.5}$$

where F^a is a spacelike vector field (on the image of its worldline) that represents the net force acting on the particle. This is the generalised form of Newton's second law for NGT.

3.2.3 Dynamically Possible Models

With these principles in mind, we are now in a position to make explicit the DPMs of NGT.[24] In NGT, one first imposes flatness of ∇_a via the field equation

$$R^a{}_{bcd} = 0. \tag{3.6}$$

A second field equation of NG is Poisson's equation,

$$h^{ab} \nabla_a \nabla_b \varphi = 4\pi\rho. \tag{3.7}$$

[24] One may question whether these laws faithfully represent Newton's thinking on these matters, since they make no reference to the persisting point of absolute space, as picked out by σ^a. For an arguably less anachronistic presentation of the laws of NGT set in Newtonian spacetime, see [55, §4.4]. The presentation of the dynamical laws of this subsection will suffice for the purposes of this paper.

Finally, the gravitational force on a point particle of mass m is given by $-mh^{ab}\nabla_b\varphi$. It follows from (3.5) that if the particle is subject to no forces except gravity, and if it has four-velocity ξ^a, then it satisfies

$$-\nabla^a\varphi = \xi^b\nabla_b\xi^a. \tag{3.8}$$

3.2.4 Symmetries of Newtonian Gravitation Theory

The above presentation of NGT in hand, consider now the symmetries of this theory. The symmetry group of NGT includes three kinds of transformations that are worth singling out: (A) the 'static shift', which involves a time-independent translation of the total matter content of the original solution; (B) the 'kinematic shift', which involves a time-independent velocity 'boost' of the total matter content of the original solution; and (C) the 'dynamic shift', which involves a time-dependent translational acceleration of the total matter content of the original solution, plus an appropriate transformation of the gravitational potential field.[25]

It is possible—and useful—to characterise all of these symmetries model-theoretically. Taking our original model to be $\mathcal{M} = \langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, \varphi, \rho\rangle$, a static-shifted model can be written $\mathcal{M}_{\mathrm{stat}} = \langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, d^*\varphi, d^*\rho\rangle$, where d is the appropriate diffeomorphism corresponding to a spatial translation. Straightforwardly—or 'literally'—understood, the world represented by $\mathcal{M}_{\mathrm{stat}}$ differs from that represented by $\mathcal{M}$ with regard to which particular points of space are underlying various parts of the matter fields. For instance, if $\mathcal{M}$ represents the centre of mass of the universe[26] as being located *here*, then $\mathcal{M}_{\mathrm{stat}}$ will represent the centre of mass of the universe as being located e.g. $3\mathrm{m}$ *to the left of here*.

Consider now the kinematic shift. The generic model yielded by applying the kinematic shift to $\mathcal{M}$ can be written $\mathcal{M}_{\mathrm{kin}} = \langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, d^*\varphi, d^*\rho\rangle$, where d is now the appropriate diffeomorphism corresponding to a velocity boost. Straightforwardly understood, the

[25] The terms 'static shift' and 'kinematic shift' are relatively standard in the literature, and are originally due to Maudlin [42, §3]. The term 'dynamic shift' is slightly less standard, and is due to Huggett [31, §8.3].

[26] One worry regarding speaking of the 'centre of mass of the universe' is the following: this notion may only be well-defined under a certain restricted set of circumstances (for example, when the mass density ρ is asymptotically zero at infinity). Given this, it may be preferable to resort to the following fix: use instead the centre of mass of some arbitrary body of matter. Our thanks to Neil Dewar for raising this point.

world represented by $\mathcal{M}_{\text{kin}}$ differs from that represented by $\mathcal{M}$ with regard to the absolute velocity of the material universe. For instance, if $\mathcal{M}$ represents the centre of mass of the universe as being *absolutely at rest*, then $\mathcal{M}_{\text{kin}}$ will represent it as moving e.g. 3ms^{-1} *due North*.

Finally, the generic model yielded by applying the dynamic shift to $\mathcal{M}$ can be written $\mathcal{M}_{\text{dyn}} = \langle M, t_{ab}, h^{ab}, \nabla_a, \sigma^a, d^*\varphi', d^*\rho \rangle$, where d is a diffeomorphism corresponding to an element of the so-called 'Maxwell group' of transformations, and where the gravitational potential field is transformed by an appropriate 'internal' transformation.[27] Thus, straightforwardly understood, the world represented by $\mathcal{M}_{\text{dyn}}$ differs from that represented by $\mathcal{M}$ with regard to what the absolute translational acceleration of the material universe is alleged to be. For instance, if $\mathcal{M}$ represents the material universe as being *absolutely non-accelerating*, then $\mathcal{M}_{\text{dyn}}$ will represent its centre of mass as accelerating in a straight line under a gravitational force-field, at e.g. 3ms^{-2} *due North*.

In sum: the symmetries of NGT include transformations that map DPMs to other DPMs that *prima facie* represent physically distinct worlds. Nevertheless, no observer 'embedded' in any of these worlds can determine which world is hers: the worlds represented by these models are 'empirically indistinguishable'—so these symmetry-related models are indeed 'empirically equivalent', in line with the definition of a symmetry transformation presented in §3.1. This is because all *relative* distances and velocities between material systems are preserved among the worlds in question, and all an observer has empirical access to are (ratios of) such distances and velocities.[28] Thus, such an observer would not be able to determine whether she is stationary, moving uniformly, or accelerating relative to the persisting points of absolute space: all of these scenarios are underdetermined by the empirical phenomena.

3.2.5 Interpretation and Motivation in Newtonian Gravitation Theory

So much for the symmetries of NGT. How do the interpretational and motivational approaches discussed in §3.1 play out in this theory? In the present context, the distinction can be stated easily. Consider again the models $\mathcal{M}$, $\mathcal{M}_{\text{stat}}$, $\mathcal{M}_{\text{kin}}$, and $\mathcal{M}_{\text{dyn}}$. According to the former

[27] Following [25, §2.3], the Maxwell group of transformations is defined as $\vec{x} \to \vec{x}' = \mathbf{R}\vec{x} + \vec{a}(t)$; $t \to t' = t + d$. The 'internal' transformation on φ is defined as $\varphi \to \varphi' = \varphi - \vec{x} \cdot \ddot{\vec{a}} + f(t)$. For further details, see [34].

[28] By 'ratios of' distances and velocities, we have in mind such notions relative to a pre-defined standard of measurement—e.g. the Parisian 'metre rod'.

view, it is legitimate to take *ab initio* all of these models—which *prima facie* represent distinct physical scenarios—to in fact represent the same state of affairs, i.e. the same possible world, even absent a coherent picture of their common ontology.

The motivational view, on the other hand, denies that it is permissible to so regard symmetry-related models as being physically equivalent. Rather, on this view, the symmetries of a theory invariably at most *motivate one to seek* a clear understanding of the common ontology underpinning such models' physical equivalence. That is, according to this view, models related by a symmetry transformation cannot be regarded as physically equivalent simpliciter. Instead, construing such models as physically equivalent is only justified once one has a clear understanding of the reality allegedly underlying them: a clear understanding that we are, according to the motivational view, invariably motivated to seek by the symmetry in question. Thus, on the motivational view, absent a clear understanding of *how it could be* that $\mathcal{M}$, $\mathcal{M}_{\mathrm{stat}}$, $\mathcal{M}_{\mathrm{kin}}$, and $\mathcal{M}_{\mathrm{dyn}}$ are to be regarded as physically equivalent (in terms of a clear explication of their common ontology), we may not regard them as so being physically equivalent, i.e. as corresponding to the same possible world.

3.2.6 Mathematical Reformulation

Importantly, the motivational approach is *not* committed to the view that whenever one is presented with a theory $\mathcal{T}$ with a symmetry between mathematically distinct models, one is motivated to *mathematically reformulate* $\mathcal{T}$ so as to remove any such (alleged) representational redundancy (such that models of the reformulated theory are constructed by *quotienting* the space of models of the original theory by the action of the symmetry in question).[29] Rather—and this will become important in our discussion of dualities in §5—we claim that such a mathematical reformulation is motivated only when the models in question are *not* isomorphic, i.e. when (straightforwardly understood) they differ more than merely with regard to which objects play which qualitative roles.[30,31] As we will now discuss, this means that, according to

[29] One may here understand 'mathematical reformulation' to mean: an alteration of the space of models of the theory (whether KPMs or DPMs). This will become clear through the examples presented in this subsection.

[30] Here, 'object' refers to any substructure of the model in question—rather than (necessarily) to the geometric objects O_i introduced in the KPMs of a generic theory $\mathcal{T}$ in §2.

[31] Note that isomorphism of two spaces of models (e.g. $\tilde{\mathcal{D}}_1$ and $\tilde{\mathcal{D}}_2$, associated respectively to two theories $\mathcal{T}_1$ and $\mathcal{T}_2$)—as introduced in §1, and discussed further in §4.1 below—should not be confused with isomorphism of a given pair of models themselves. It is the latter that is under consideration here.

the motivational view, in the case of NGT only the kinematic and dynamic shifts motivate us to mathematically reformulate the theory so as to remove any representational redundancy.

The static shift in NGT is crucially distinct from the kinematic and dynamic shift—for the reason that the models in question in this case are isomorphic. And indeed, there exists a straightforward means of understanding such isomorphic models' physical equivalence, which necessitates no mathematical reformulation of the theory. This view goes by a variety of names in the literature: in spacetime contexts, it is most commonly referred to as 'sophisticated substantivalism'.[32] The sophisticated substantivalist denies that spacetime points possess primitive transworld identities; instead, they are 'contextually individuated' [35, §5]: they are not to be construed as being anything less, or more, than 'nodes' in the relational, geometrical structures in which they are embedded. This view is still a version of substantivalism, in the sense that it is committed to points of space being fundamental, basic elements of reality. Crucially, however, this view denies that there are any primitive, singular (haecceitistic) facts about spacetime points (e.g. *this particular* point of space is materially occupied) which would even allow for a physical distinction between statically shifted scenarios to be drawn.

Analogous considerations apply in the context of more modern physical theories; the diffeomorphism invariance of GR provides a case in point. Just as for the static shift in NGT, the existence of this symmetry is alleged to commit the substantivalist to a plurality of physically distinct, but nevertheless empirically indistinguishable, possibilities. Once again, we can phrase this in model-theoretic terms: taking a generic DPM of GR, $\mathcal{M} = \langle M, g_{ab}, \Phi \rangle$, we can apply an arbitrary diffeomorphism d to yield a new DPM, $\mathcal{M}' = \langle M, d^* g_{ab}, d^* \Phi \rangle$. A popular allegation—the canonical version of which can be found in [26, §4]—is that the spacetime substantivalist is committed to regarding the two worlds represented by these models as differing with regard to which particular points of the spacetime manifold are underlying various parts of the metric and matter fields.[33]

We hope it is clear that, if adopting sophisticated substantivalism constitutes a legitimate response to the alleged problem of NGT's static shift symmetry, it should count as an equally

[32] See, e.g., [53, p. 575]. Other names for this view include 'moderate structural realism' about spacetime [27, pp. 31-2] and 'non-reductive relationalism' [64, §5].

[33] Here we ignore the related (but distinct) 'indeterminism' objection to substantivalism in the context of GR raised at [26, §5]. The reasons for this are twofold. First, this objection is not directly related to the static shift argument in NGT. Second, sophisticated substantivalism also seems sufficient as a response (for more on this latter point, see [51, §4.1.4]).

legitimate response to the alleged problem of GR's diffeomorphism symmetry. That is, adopting sophisticated substantivalism should be sufficient for one to be able to understand, in a perfectly transparent way, how it is that diffeormorphism-related models in GR are to be regarded as physically equivalent, without any mathematical reformulation of the theory being necessitated—just as in the case of the shift symmetry of NGT.[34]

Now return to the case of non-isomorphic symmetry-related models in NGT, namely $\mathcal{M}$, $\mathcal{M}_{\text{kin}}$, and $\mathcal{M}_{\text{dyn}}$. 'Literally understood', such models do not represent possible worlds which differ merely haecceitistically. Hence, adopting sophisticated substantivalism is by itself insufficient to be able to understand how such models are to be regarded as physically equivalent.[35] Thus, (we claim) we are motivated to mathematically reformulate the theory so as to obtain a coherent understanding of the common ontology underpinning such models' physical equivalence.[36]

Such a mathematical reformulation of the theory is indeed possible. In fact, for the kinematic shift, it is trivial: one simply excises σ^a from KPMs of the theory—so that the question of two otherwise-identical models differing only in the absolute velocity of the centre of mass of the matter content they represent does not arise. In other words, one moves from Newtonian spacetime—where the persistence of points of space through time is assured (since, recall, the histories of these points are associated with the integral curves of the σ^a field), and where the associated notion of absolute velocity is physically meaningful—to Galilean spacetime, where the persistence of points of space through time and the associated notion of absolute velocity no longer make physical sense, but where the difference between straight (inertial) and curved (accelerating) trajectories through spacetime remains physically meaningful.[37]

In the case of the dynamic shift, reformulation is also possible—though somewhat less

[34]Cf. footnote 15. Of course, sophisticated substantivalism constitutes just one of many positions available in the vicinity of discussions of the hole argument. For a recent review of the literature, see [54, §7].

[35]For the parallel point in the case of dualities, see [59, §5.3]. Cf. §5.1.

[36]Note that, since the mathematically reformulated theory will have a different space of models to the original theory (cf. footnote 29), it may best be regarded as a *new* theory, distinct from the original (on the setup of §2).

[37]For further discussion, see e.g. [25, §2.4] and [43, pp. 54-66]. Although such a reformulation of NGT may appear trivial from a modern four-dimensional, differentio-geometric perspective, it certainly would not have appeared so to Newton or his contemporaries. This appearance of triviality is arguably reinforced by the fact that, in setting up NGT, we have (following the canonical literature on this subject, in particular [28, pp. 71-94]) formulated the laws directly in terms of ∇_a, rather than σ^a. For more on this point, see [55, p. 134]; for a discussion of NGT which puts particular emphasis on the non-triviality of the move to Galilean spacetime, see [43, pp. 54-66].

straightforward. (Those readers uninterested in the technical details here are advised to skip both this and the following paragraph.) Here—having already eliminated σ^a from the models of the theory—one replaces the flat derivative operator ∇_a of NGT with a (partly) dynamical derivative operator $\hat{\nabla}_a$,[38,39] for which the DPMs then require that the associated curvature tensor $\hat{R}^a{}_{bcd}$ satisfies

$$\hat{R}_{bc} = 4\pi\rho t_b t_c, \tag{3.9}$$

$$\hat{R}^a{}_b{}^c{}_d = \hat{R}^c{}_d{}^a{}_b, \tag{3.10}$$

$$\hat{R}^{ab}{}_{cd} = 0, \tag{3.11}$$

and one also eliminates the gravitational potential φ from KPMs of the theory—so that they are quintuples $\langle M, t_{ab}, h^{ab}, \hat{\nabla}_a, \rho\rangle$. (3.9) is the geometrised version of Poisson's equation (3.7); (3.10) holds in a classical spacetime iff this admits, at least locally, a smooth, unit timelike field ξ^a that is geodesic ($\xi^b\nabla_b\xi^a = 0$) and twist-free ($\nabla^{[a}\xi^{b]} = 0$) [36, p. 281]; (3.11) holds throughout M iff parallel transport of spacelike vectors in M is, at least locally, path-independent [36, p. 279]. The resulting theory is known as 'Newton-Cartan theory' (NCT).

It can be shown—via Trautman's geometrisation and recovery theorems[40]—that the class of models of NGT which differ by a dynamic shift all map (up to isomorphism) to the same model of NCT. Moreover, for all timelike curves of M with four-velocity field ξ^a, particles subject to a gravitational force in NGT (so that $\xi^b\nabla_b\xi^a = -\nabla^a\varphi$) move along geodesics in NCT (so that $\xi^b\hat{\nabla}_b\xi^a = 0$). In other words, in NCT gravity is no longer a force, as in NGT.

For our purposes, the crucial point to note is the following: by reformulating NGT *à la* NCT as per the above, one constructs a mathematical reformulation of the theory which eliminates the gauge redundancy (in the sense of §2) manifest in the possibility of a dynamic shift in

[38] $\hat{\nabla}_a$ is related to ∇_a by $\hat{\nabla}_a = (\nabla_a, C^a{}_{bc})$, with $C^a{}_{bc} = -t_b t_c \nabla^a\varphi$; bracket notation for derivative operators means that $\hat{\nabla}_a$, ∇_a, and $C^a{}_{bc}$ are related by $(\nabla'_c - \nabla_c)\,\alpha^{a_1\ldots a_r}_{b_1\ldots b_s} = \alpha^{a_1\ldots a_r}_{db_2\ldots b_s} C^d{}_{cb_1} + \ldots + \alpha^{a_1\ldots a_r}_{b_1\ldots b_{s-1}d} C^d{}_{cb_s} - \alpha^{da_2\ldots a_r}_{b_1\ldots b_s} C^{a_1}{}_{cd} - \ldots - \alpha^{a_1\ldots a_{r-1}d}_{b_1\ldots b_s} C^{a_r}{}_{cd}$; and t_a is a covector field which may (locally) be defined from t_{ab} via $t_{ab} = t_a t_b$ in a 'temporally orientable' spacetime—for details, see [36, pp. 250-251].

[39] We say 'partly' rather than 'fully' dynamical in light of the compatibility conditions (3.2) and (3.3), which hold also for $\hat{\nabla}_a$.

[40] For original sources, see [70]; for contemporary discussion and proofs, see [36, pp. 267ff.].

NGT.[41] Note also the important point that moving to NCT is not by itself sufficient to be able to understand as physically equivalent all symmetry-related models of Newtonian theory set in flat spacetime. This is because—as mentioned in the previous paragraph—such symmetry-related models will typically correspond to a single model of NCT only up to isomorphism. Thus, in order to have a fully transparent understanding of how it is that symmetry-related models of Newtonian theory set in flat spacetime can correspond to a single model of NCT, a sophisticated substantivalist conception of spacetime ontology is also required.[42]

To summarise: According to the interpretational approach, it is typically legitimate *ab initio* to regard symmetry-related models as being physically equivalent, even absent a coherent explication of their common ontology. According to the motivational approach, by contrast, it is not legitimate *ab initio* to regard symmetry-related models as being physically equivalent. Rather, symmetries invariably at most *motivate one to seek* a coherent explication of the common ontology underpinning such models' physical equivalence. When the symmetry-related models in question are not isomorphic, one is motivated to mathematically reformulate the theory. On the other hand, when they are isomorphic, one is not motivated to mathematically reformulate the theory: adopting 'moderate structuralism'—which, in the spacetime context, we take to be equivalent to sophisticated substantivalism[43]—is invariably sufficient.

3.3 Motivating Motivation

Up to this point, we have remained officially neutral between the interpretational and motivational approaches to symmetry transformations. Here, however, developing upon [45, §4], we wish to argue explicitly for the latter. One argument in favour of this position is the following: even if the central claim of the interpretational approach—that one may legitimately

[41] For further details, see [34]. Note also that Saunders [65] has argued on the basis of similar considerations that the appropriate spacetime setting for Newtonian theory is in fact 'Newton-Huygens spacetime', a close relative of what Earman [25, §2.3] has dubbed 'Maxwellian spacetime'. For illuminating discussion of Saunders' paper, see [22, 69, 75, 77].

[42] A similar moral applies in the case of moving to Galilean spacetime as a response to NGT's boost invariance. Many thanks to David Wallace for pushing us on this point.

[43] We draw the term 'moderate structuralism' from [27]; compare also the 'modest structuralism' of [52, p. 102]. According to this view, objects (e.g. points of spacetime) are construed as being nothing more (or less) than 'nodes' in the relational structures in which they are embedded; and the possibility of purely haecceitistic distinctions between worlds is denied. Construed in this way, moderate structuralism encompasses sophisticated substantivalism—but is a stronger thesis due to the latter clause.

regard certain symmetry-related models of a theory as being physically equivalent even in the absence of a coherent picture of their common ontology—is true, on this approach, the *reality in terms of which* this physical equivalence is to be understood will, absent further details, remain opaque. That is, without further work, the advocate of the interpretational approach offers no explanation as to how such physical equivalence is to be construed, or how it could even be said to arise. To the extent that the interpretational view is not supposed to reduce to an uninteresting form of instrumentalism, it is unclear what realistic picture of the world is being propounded by the defender of this position; it is opaque what, according to her, *the world really is like*.

Here, one must separate two closely-related points. First, the advocate of the interpretational approach may be regarded as shirking her responsibility to provide a coherent explication of the common ontology associated with symmetry-related models; this, however, is where much of the most interesting work in the foundations of physics is done.[44] Second, the advocate of the interpretational approach makes at the outset an assumption that certain symmetry-related models admit of a coherent interpretation which makes manifest their physical equivalence.[45] In our view, it is more cautious to avoid such an assumption: to only regard symmetry-related models as being physically equivalent once an explication of their common ontology can be provided; to consider us always motivated to attempt to construct such an explication; and thereby to favour the motivational over the interpretational approach.

Having said this, it is worth distinguishing two sub-views within the motivational approach. According to the former, more *confident* view, symmetry-related models may only be regarded as being physically equivalent once an interpretation affording a coherent explication of their common ontology is provided, *but such an interpretation is always guaranteed to exist*. By contrast, according to the latter, more *cautious* view, symmetry-related models may only be regarded as being physically equivalent once a coherent explication of their common ontology is provided, *and there is no guarantee that such an interpretation exists*. It should be clear from the foregoing that we favour the latter, more cautious strand of motiva-

[44]For example, the staunchest advocate of the interpretational approach would likely not be motivated to consider whether NGT can be reformulated in terms of Galilean spacetime, or NCT: the bare assertion that models of NGT related by kinematic or dynamic shifts are physically equivalent effectively eliminates motivation for the advocate of the interpretational approach to pursue this research programme.

[45]Cf. footnote 14. Here, the advocate of the interpretational approach may be guided by overarching, *a priori* principles, connecting certain features of the symmetry-related models under consideration with their physical equivalence. However, unless a necessary connection between such features and the physical equivalence of the models can be forged, the point in the body of this paragraph stands.

tionalism.[46] Articulating these two distinct versions of the motivational approach will prove illuminating, when it comes to constructing a taxonomy of the views of philosophical authors in the parallel case of dualities—cf. in particular §5.3.

4 Equivalence and Duality

In the previous section, we distinguished the interpretational and motivational approaches to symmetry transformations, and defended the latter, expanding upon [45, §4]. In this section, we introduce the notions of 'theoretical equivalence' and 'underdetermination', both of which will prove important in our defence of the motivational approach to dualities in §5.

4.1 Theoretical Equivalence and Duality

We now introduce a notion of 'theoretical equivalence'. Given two theories $\mathcal{T}_1$ and $\mathcal{T}_2$, with respective spaces of DPMs $\mathcal{D}_1 \subset \mathcal{K}_1$ and $\mathcal{D}_2 \subset \mathcal{K}_2$, we say (broadly following [76,78,79]) that these are 'theoretically equivalent' iff (i) there exists an isomorphism between $\tilde{\mathcal{D}}_1$ of $\mathcal{T}_1$ and $\tilde{\mathcal{D}}_2$ of $\mathcal{T}_2$;[47] and (ii) the empirical predictions corresponding to each $\tilde{\mathcal{M}}_1 \in \tilde{\mathcal{D}}_1$ are identical to the empirical predictions corresponding to the associated $\tilde{\mathcal{M}}_2 \in \tilde{\mathcal{D}}_2$.[48] If (i) holds of two theories but not (ii), then we say that they are (merely) 'formally equivalent'; if (ii) holds of two theories but not (i), then we say that they are (merely) 'empirically equivalent'.[49] Two theories are theoretically equivalent iff they are formally equivalent and empirically equivalent.

Turn now to the notion of duality: a pervasive phenomenon in string theory.[50] The four

[46] We owe the nomenclature of 'confident' versus 'cautious' versions of the motivational approach to Jeremy Butterfield.

[47] Recall from §2 that, for a given theory $\mathcal{T}$, $\tilde{\mathcal{D}}$ denotes the gauge-reduced space of DPMs of $\mathcal{T}$. Note also that one may introduce a graded notion of theoretical equivalence, by imposing restrictions on the structure of the models preserved by this isomorphism. Though important to note, this latter point will be set aside in this paper.

[48] Each $\tilde{\mathcal{M}}_1 \in \tilde{\mathcal{D}}_1$ and $\tilde{\mathcal{M}}_2 \in \tilde{\mathcal{D}}_2$ which correspond to the same empirical predictions in this way may be said to be 'empirically equivalent'—cf. §3.1. Note that two models may be empirically equivalent without the theories to which they belong being empirically equivalent, in the sense given below.

[49] The map between $\tilde{\mathcal{M}}_1 \in \tilde{\mathcal{D}}_1$ and $\tilde{\mathcal{M}}_2 \in \tilde{\mathcal{D}}_2$ may not be one-one in the absence of formal equivalence.

[50] For philosophically-oriented introductions to dualities—including all the theories and their respective dualities mentioned in this paragraph—see e.g. [50, 60]. For more recent and nuanced philosophical approaches to dualities, see [9, 15, 17].

best-known examples of dualities arising in string theory are 'T-duality', 'mirror symmetry', 'S-duality', and the 'AdS/CFT correspondence'. In the case of T-duality, type IIA superstring theory on a product manifold $M \times S^1$ with radius of the periodic dimension R is found to be theoretically equivalent to type IIB superstring theory on the product $M \times S^1$ with radius of the periodic dimension proportional to $1/R$ [4, ch. 6]. Mirror symmetry is a generalisation of T-duality to the case of topologically inequivalent manifolds.[51] S-duality relates models of one superstring theory with string coupling constant g_s to models of another superstring theory with string coupling constant $1/g_s$; it is thus a so-called 'strong/weak' duality. For example, strongly/weakly coupled type I superstring theory is theoretically equivalent under S-duality to weakly/strongly coupled $SO\,(32)$ heterotic string theory [4, §8.2]. Finally, in the AdS/CFT correspondence (originally introduced in [37]), a string theory in so-called 'AdS spacetime' is theoretically equivalent to a conformal field theory (CFT) in a lower number of spacetime dimensions [4, ch. 12].

In addition to being theoretically equivalent (indeed, we take theoretical equivalence as the *definition* of a duality in the ensuing), should duality-related models also be understood as being physically equivalent—i.e., as representing the same possible world? Before we attempt to answer this question, two related points are worth stating. First, dualities are (more) analogous to the kinematic and dynamic shifts than to the static shifts of §3.2.4—for the reason that dual models are generically not isomorphic: straightforwardly understood, they represent worlds which differ more than purely with regard to which particular objects are playing which qualitative roles (see e.g. [59, p. 224]). Hence—as we will elaborate—dualities in general motivate mathematical reformulation; adopting moderate structuralism is by itself insufficient to understand satisfactorily the (alleged) physical equivalence of duality-related models. Second, the apparent physical difference between duality-related models can be much more striking than in the case of e.g. models of NGT related by kinematic and dynamic shifts: naïvely understood, the possible worlds they represent are *very* different. For instance, models related by an AdS/CFT-type duality differ in the number of dimensions they (appear to) attribute to spacetime; models related by mirror symmetry differ in the topology they (appear to) attribute to spacetime.

[51] For a philosophical introduction to mirror symmetry, see [62].

4.2 Underdetermination

The distinction between formal and empirical equivalence is of value when discussing whether a pair of theories exhibits 'strong underdetermination of theory by evidence'. We say that two theories $\mathcal{T}_1$ and $\mathcal{T}_2$ present such a case when they are empirically equivalent, yet there exists at least one pair of empirically equivalent models $\tilde{\mathcal{M}}_1 \in \tilde{\mathcal{D}}_1$ and $\tilde{\mathcal{M}}_2 \in \tilde{\mathcal{D}}_2$ which are nevertheless interpreted as corresponding to distinct possible worlds, respectively W_1 and W_2.

To have a case of strong underdetermination, the two theories must be empirically equivalent; however, one may ask whether formal equivalence is also relevant. Distinguish:

(A) $\mathcal{T}_1$ and $\mathcal{T}_2$ being empirically equivalent *and* formally equivalent.

(B) $\mathcal{T}_1$ and $\mathcal{T}_2$ being empirically equivalent *but not* formally equivalent.

Any argument to the effect that instances of (A) cannot lead to strong underdetermination[52] is (roughly) in line with the Quinean position according to which theories related by reconstrual of predicates (the analogue of formal equivalence) are understood not to lead to such underdetermination [57].[53,54] However, consider again models related by e.g. the AdS/CFT correspondence, or mirror symmetry. In spite of an instantiation of formal equivalence, such models of these theories at least *appear* to be ontologically distinct, and thus to correspond to distinct possible worlds.[55] Now, when it comes to deciding whether (A) can indeed lead to strong underdetermination, one can be aided by one's prior commitments in the philosophy of science: even if one is a realist, it may be that by e.g. giving some structuralist account, one can make plausible that models of such theories correspond to the same world (in this regard

[52] I.e. to the effect that each $\tilde{\mathcal{M}}_1 \in \tilde{\mathcal{D}}_1$ and its associated $\tilde{\mathcal{M}}_2 \in \tilde{\mathcal{D}}_2$ must correspond to the same world.

[53] For a critical discussion of the Quinean approach to theoretical equivalence, see [1]. With the conclusion of that paper—"If one takes Quine equivalence as the standard for theoretical equivalence, one underestimates the threat of underdetermination" [1, p. 483]—we are in agreement. Cf. also [2].

[54] To allay any possible misunderstanding: Quine's view on this matter is *not* that theories with isomorphic spaces of solutions are always theoretically, or even empirically, equivalent. Rather, Quine holds a strictly stronger view on what it is for two theories to be theoretically equivalent: *if* theories are related by a suitable reconstrual of predicates, *then* such theories are theoretically equivalent. This, plausibly, *entails* that their respective spaces of solutions are isomorphic. But, for Quine, the fact that theories have isomorphic spaces of solutions does not by itself entail that they are theoretically or even empirically equivalent.

[55] The concern, therefore, is over the adequacy of formal equivalence—which is, indeed, a formal notion—to capture an informal or semantic notion: that of two models representing the same possible world. Our thanks to Jeremy Butterfield for suggesting that we put the matter in this way.

in the context of dualities, see [39]).[56,57] Nonetheless, examples such as this demonstrate that instances of (A) might, *prima facie*, give rise to strong underdetermination (cf. [39, p. 474])—*pace* Quine, who argued in [57] that only instances of (B) could constitute genuine cases of strong underdetermination.[58] The relevance of this for our views on the interpretation of dualities will become apparent in §5.

5 Duality as Motivation

5.1 Interpretation and Motivation, Reprise

Above, we saw the *prima facie* plausibility of interpreting duality-related models as representing distinct possible worlds. This, however, runs against a common view, that duality-related models may typically be taken *ab initio* to represent the same possible world. For example, Rickles writes:[59]

> [D]ual theories are simply examples of theoretically equivalent descriptions of the same underlying physical content: I distinguish them from cases of genuine underdetermination on the grounds that there is no real incompatibility involved between the descriptions. The incompatibility is at the level of purely unphysical structure. I argue that dual pairs are in fact very strongly analogous to gauge-related solutions ... I conjecture that dualities always point to a more fundamental

[56] This said, Rickles has recently suggested that cases such as the AdS/CFT correspondence may give rise to *structural* underdetermination [60, 61, 63]. (We concur with this view; cf. footnote 57 below.) Note that if this is so, then even the structuralist may not be able to argue that such pairs of formally equivalent models correspond to the same possible world.

[57] In our view, adopting structural realism as a means of identifying symmetry-related models succeeds only if the models in question are 'naïvely' understood as representing at most haecceitistically distinct possible worlds. In that case, it is clear how adopting structural realism allows us to identify such (putatively) distinct physically possibilities as (actually) not distinct after all. However, if the models in question are 'naïvely' understood as representing more than haecceitistically distinct possible worlds, then adopting structural realism (by itself) is insufficient to provide grounds for understanding the models in question as corresponding to the same possible world.

[58] If this is correct, then we concur with De Haro *et al.* that "we will need to allow that formal isomorphisms do not in general imply sameness of content" [19, §3.1]. Cf. footnote 55.

[59] For further clear expression of this position, see e.g. [40, 41].

(intrinsic) description, namely that in which the representational redundancy is eliminated. [63, p. 62]

This position bears striking similarity to the interpretational approach to symmetries.[60] Indeed, by analogy, we may define at this juncture an interpretational approach to dualities. According to this, when presented with a pair of duality-related theories, we are (a) typically warranted in first interpreting duality-related models as representing the same possible world—even absent a coherent explication of their common ontology;[61] then may (but are not required to) go on to (b) identify those pairs of dual models, thereby constructing the space of KPMs $\bar{\mathcal{K}}$ of a new theory $\bar{\mathcal{T}}$, which represents the 'common core' (in the language of [15, §2.2]) of the duality-related theories; and (c) having constructed such a $\bar{\mathcal{K}}$, seek to provide a coherent picture of the ontology of the models $\bar{\mathcal{M}} \in \bar{\mathcal{K}}$.

In contrast with this interpretational approach to dualities, one may also define a motivational approach to dualities. According to this view, the existence of a duality between two theories first (a) motivates us to provide a coherent picture of the common ontology of the pairs of models of these two theories related by the duality; but only once such a characterisation is constructed should we (b) interpret those models as representing the same possible world; and (potentially) (c) identity those models to construct a space of KPMs $\bar{\mathcal{K}}$ of some new theory $\bar{\mathcal{T}}$, which represents the 'common core' of the duality-related theories.

It is important to distinguish two sub-positions within the interpretational view. On the first such view, the existence of a duality motivates us to seek a clear picture of the ontology alleged to underlie the dual-related models.[62] On the second view—by contrast—no such search for a coherent understanding of the common ontology of the dual models is required.[63] Importantly, however (and to reiterate), both of these sub-positions are consistent with the interpretational view. By contrast, the advocate of the motivational approach to dualities maintains that it is *only* legitimate to regard duality-related models as being physically equivalent *if* one possesses a clear picture of the common ontology of the dual models.

[60] One might argue that the final sentence here is in line with the motivational approach. Even if this is true, however, in our view it is not correct to read Rickles as *endorsing* the motivational approach, in light of the preceding sentences in the quote—see below.

[61] Here, the same points from §3.1 regarding the 'typicality' clause arise again.

[62] This view appears more popular in the literature; cf. again e.g. [63, p. 62], and footnote 60.

[63] This can be considered the analogy of Dewar's approach to symmetries in the case of dualities [20, p. 322].

Following §3.3, one can also introduce two sub-positions within the motivational approach to dualities—according to the former, more *confident* view, duality-related models should not be regarded as being physically equivalent absent a coherent explication of their common ontology, but such an explication is always guaranteed to be found; according to the latter, more *cautious* view, duality-related models should not be regarded as being physically equivalent absent a coherent explication of their common ontology, and no such explication is guaranteed to be found. This distinction between sub-positions within the motivational approach will prove to be illuminating in §5.3, when we consider the views of certain authors *vis-à-vis* the interpretation of duality-related models.

Regardless of where they may stand in the above debate, many authors—both physicists and philosophers—maintain that string-theoretic dualities motivate us to seek a mathematical reformulation of the dual string theories.[64] Since duality-related models are generically non-isomorphic, we concur with this verdict (cf. §4.1). Where the advocate of the motivational approach disagrees with some of such authors, however, is on the question of whether, *in the absence of any mathematical reformulation of string theory*, it is legitimate to regard duality-related models as being physically equivalent: in her view, it is not. Moreover, it is precisely on this issue that we disagree with the commonly-held interpretational view.[65]

Given the importance of this point for the purposes of this paper, it is worth repeating. When presented with two symmetry- or duality-related models, the interpretationalist will

[64] Or quantum field theories, in the case of e.g. the AdS/CFT correspondence.

[65] Perhaps it is true that many recent philosophical authors' views on dualities are more subtle than a straightforward endorsement of the interpretational approach, *à la* Rickles [63, p. 62]. This notwithstanding, however, a reader unfamiliar with the literature on dualities may obtain the *impression* that the interpretational approach is widely embraced. Here is some *prima facie* evidence to support this claim: "[D]ual [theories] should be understood as giving physically equivalent descriptions" [32, pp. 87-88]; "In all dualities, it is the theories that are equivalent. ... [C]ertain transformations 'don't matter'. The only difference between these [dual] transformations and standard gauge symmetries is that they seem to relate things that *look like they really should matter*!" [63, p. 64]; "[Our] conception of duality meshes with two dual theories being 'gauge-related', in the general philosophical sense of being physically equivalent. For a string duality, such as T-duality and gauge/gravity duality, this means taking such features as the radius of a compact dimension, and the dimensionality of spacetime, to be 'gauge"' [19, p. 68]; "The stance adopted [in this paper] is ... to avoid a literal reading of the elementary/composite interchange and, on this basis, to avoid mixing the question of its meaning with the question of physical fundamentality. The attitude is analogous to the one shared in this volume [a recent special issue of *Studies in the History and Philosophy of Modern Physics* devoted to dualities] about how to understand apparently puzzling features such as the interchange of tiny and huge dimensions connected with T-duality in string theory, or the duality of dimension under the AdS/CFT (gauge/gravity) correspondence. The underlying idea is that, what the dual descriptions do not agree upon, should not be attributed a real physical significance. In fact, this means nothing else than saying that the physics (including its ontology) remains the same under the duality. What changes, is just the way of looking at it" [10, p. 101].

typically say that it is legitimate to regard such models as physically equivalent. The motivationalist will deny this: for her, the mere fact that models are related by a symmetry or duality transformation is not a sufficient reason to regard them as physically equivalent. *This* is what crucially separates the interpretationalist and motivationalist positions.

The motivationalist will go on to say that, for any two symmetry- or duality-related models, we are motivated to try to provide an explication of the common ontology that is alleged to underlie them. *The interpretationalist will not always agree*. Some interpretationalists (e.g. Rickles [63]) *will* claim that that we are so motivated. But not all will (e.g. Dewar [20]). In other words, the interpretationalist and the motivationalist *do not invariably agree about motivation*. Thus, *merely* claiming that dualities motivate us to formulate (e.g.) 'M-theory'[66]—which is conjectured to be the theory which would transparently explain dual string theories' physical equivalence—is not sufficient to make one a motivationalist.[67]

To close this subsection, it is worth reflecting further on the nature of the ontology 'common' to two dual models; two broad attitudes are possible here. First, given two theories understood to be dual, one may attempt to identify the 'shared structure' across duality-related models;[68] one may then use this as a guide to the interpretation of the dual theories. This austere approach to the interpretation of dualities is advanced in e.g. [15, §2.2]. On the other hand, one may be more ambitious. For example, one may construct a new theory, such that dual models of the original theories each constitute (partial) descriptions of certain models of the new theory. The approach of attempting to find an overarching 'M-theory', of which all five superstring theories are 'limits' (in some appropriate sense—cf. §5.2), fits naturally into this latter category. It is worth remarking, however, that as it stands the existence of such a theory—as well as our ability to discover it—remains conjectural; for further discussion,

[66] See below for further discussion of this theory.

[67] An anonymous reviewer has questioned whether our nomenclature for these two positions is entirely apt. This is for two main reasons: (i) It appears to obscure the fact that *both* the interpretational and motivational approaches are, broadly speaking, possible ways of 'interpreting' theories; (ii) It might suggest that the interpretationalist *never* agrees with the motivationalist on the issue of whether one is motivated to reformulate a given theory in the presence of certain symmetries. With regard to (i): We agree that both approaches are, in some sense, possible ways of 'interpreting' physical theories, although we disagree that our nomenclature truly obscures this fact. (Our claim is simply that someone who endorses the interpretational approach to symmetries is not 'interpreting' symmetries the right way; the right way to 'interpret' them is motivationally!) With regard to (ii): As noted in the main text, although the interpretationalist *might* agree with the motivationalist on the issue of motivation in certain cases, she need not always do so. Motivation is *not* an essential component of the interpretationalist's position, as it is for the motivationalist.

[68] That is, the formal structure preserved across duality-related models—cf. [15, 21]. This is the 'common core' of the two dual models, in the sense given above.

see [60, §5.2] and [59, §5]. For a recent, detailed clarification of the distinction between these two approaches to explicating the ontology 'common' to duality-related models, see [16].

5.2 Motivating Dualities as Motivation

As with the debate between advocates of the interpretational and motivational approaches to symmetries (cf. §3.3), we endorse the latter approach to dualities over the former. Our reasons for doing so broadly mirror those given in §3.3. First, just as in the case of symmetry transformations, the interpretationalist may be regarded as shirking her responsibility to provide a coherent explication of the common ontology associated to duality-related models. Second, the interpretationalist assumes at the outset that certain duality-related models admit of a coherent interpretation which makes manifest their physical equivalence. In our view, however, it is more cautious to drop such an assumption: to only regard duality-related models as being physically equivalent once an explication of their common ontology can be provided; to consider us always motivated to attempt to construct such an explication; and thereby to favour the motivational over the interpretational approach to dualities. (We return in a moment to the distinction between two strands of motivationalism drawn in §§3.3 and 5.1.)

With the above in mind, recall now that some authors, such as Polchinski, *define* duality such that "we have a single quantum system that has two classical limits" [50, p. 7]—from which one concludes that "it is fruitless to argue whether [$\mathcal{T}$] or [$\mathcal{T}'$] provide [*sic*] the fundamental description of the world; rather, it is the full quantum theory" [50, p. 7].[69] If one approaches dualities in this manner, then one begins with a single theory—models of which may be interpreted as corresponding to certain possible worlds—and constructs two further (dual) theories therefrom. In that case, one is already in possession of a coherent account of the physical equivalence of models of the two dual theories, in terms of the ontology of models of the 'quantum' theory from which they are constructed.

If one follows this approach, then one may argue that the interpretational account of dualities is favoured over the motivational—for one may indeed interpret duality-related models as being physically equivalent *ab initio*, via an explication of the ontology of their underly-

[69] Such an approach to dualities is also implicit in [40,41]. In these passages, by a "quantum system", Polchinski means a quantum theory; by two "classical limits", he means two theories for which perturbation theory is applicable, which may be defined from the original quantum theory—see [50, pp. 6ff.].

ing ‘quantum’ theory. Note, however, that this is also consistent with the motivational account: since such an explication can be provided in every case, motivation for regarding the duality-related models as being physically equivalent is automatically secured—so we are indeed warranted in doing so.

One may, however, question whether Polchinski’s account of dualities is most appropriate—for it is not the case that we always construct dual theories from such an underlying theory. For example, in the case of the AdS/CFT correspondence, neither the string theory in anti-de Sitter space, nor the boundary conformal field theory, was constructed from a third, ‘quantum’ theory. Indeed, this is also true in the case of string-theoretic dualities: though certain dual string theories are conjectured to be limits of an underlying so-called M-theory (introduced by Witten in [81]), the existence of this theory was postulated *post facto*[70]—and it is not the case that these string theories were initially defined therefrom. Nevertheless, there is overwhelming evidence (in terms of matching of correlation functions, etc.; see e.g. [4,18]) that such theories are dual, in the sense of §4.1.[71]

It is these cases—in which a duality between theories is discovered after the fact, and the existence of an underlying, third theory is only conjectured—which are most interesting from the point of view of the interpretational/motivational distinction. In such cases, we take it that one needs to give a coherent account of the shared ontology (i.e. an appropriate interpretation) of duality-related models, before one declares those models to be physically equivalent; moreover, there appears to exist no set of *a priori* principles by which one may deductively infer that such an interpretation exists. For these reasons, we believe that (a) the definition of duality presented in this paper is broader, more flexible, and more faithful to string theory history than that proposed by Polchinski; and (b) that the right approach to such dualities is the motivational approach.[72]

[70]The *facto* here being the construction of the two original, dual theories.

[71]Specifically, *vis-à-vis* both their theoretical equivalence and their empirical equivalence.

[72]If one endorses these views, then one will argue that more needs to be done to demonstrate the physical equivalence of duality-related models of certain string theories—in terms of exploring the mathematics and interpretation of M-theory—before such physical equivalence can be declared by appeal to this (conjectured) theory.

5.3 Confident and Cautious Motivational Approaches

Other authors working in the foundations of quantum gravity—most notably De Haro—also appear to endorse something akin to the motivational approach to dualities:

> Duality in mathematics is a formal phenomenon: it does not deal with physically interpreted structures ... But this is also how the term is used by physicists: it is attached to the equivalence of the formal structures of the theories, regardless of their interpretations, i.e. without it necessarily implying the physical equivalence of the theories which describe two concrete systems. ...
>
> Duality, then, is one of the ways in which two theories can be theoretically equivalent, without its automatically implying their physical equivalence. [15, p. 9]

We concur with this verdict, though it is worth exploring further De Haro's position. To do so, first follow De Haro in distinguishing (i) 'extendable' and 'unextendable' theories—i.e. "theories which do, respectively do not, admit suitable extensions in their domains of applicability" [15, p. 4]—and (ii) 'external' and 'internal' interpretations of theories—i.e. "interpretations which are obtained from outside (i.e. by coupling the theory to a second theory which has already been interpreted), respectively from inside, the theory" [15, p. 4].[73]

De Haro argues that, since extendable theories may be coupled to further theories, they should *not* be regarded as being physically equivalent.[74] His reasoning here is motivated by Galileo ship-type scenarios. To see this, consider the same physical system (e.g. the ship), coupled to two further *different* physical systems (e.g. the shore at rest versus the shore in motion)—we should (De Haro claims) not necessarily regard these two extendable models as being physically equivalent *tout court*, in light of possible couplings to other physical systems which suffice to reveal their physical distinctness. The same point applies to *dual* extendable theories: though they are intertranslatable via the formal duality map, they should *not* be regarded as being physically equivalent, in light of possible further couplings to other physical systems.

[73] For further details, see [15, §1]. Cf. also [21], in which a very similar distinction is drawn.

[74] For further discussion of these matters in the symmetries literature, see [29, 68]; cf. also footnote 16.

Though we concur with De Haro on this point—indeed, ultimately we *fully* agree with De Haro (and his close collaborator, Jeremy Butterfield) on these matters—it is worth reflecting further upon his writings (and those of Butterfield), since doing so brings to light two different sub-views within the motivational approach (*viz*., the *confident* and *cautious* versions of the motivational approach, discussed in §§3.3 and 5.1). To this end, consider the following passage from Butterfield: [9, p. 5]

> ... De Haro proposes a sufficient condition for one to be justified in interpreting two duals to be physically equivalent. This sufficient condition has two conjuncts. The first is that each dual is *unextendable*: which means, roughly speaking, that the dual, i.e. the theory, both is a complete description of its intended domain and cannot be extended to a larger domain. The second is that each dual has an *internal* (as against *external*) interpretation: i.e. an interpretation that does *not* proceed by coupling to another theory, often one which describes measurements of the given theory's domain. These conjuncts are linked in that De Haro argues that unextendability implies that one is justified in using an internal interpretation: (justified but not obliged—there can still be external interpretations).

We fully agree with this passage. Indeed, Butterfield goes on to write that "a statement of the bare theory, and an internal interpretation, are not automatic, given a proven duality" [9, p. 43], and to argue that (e.g.) Newton and Clarke were right not to move from Newtonian mechanics formulated in Newtonian spacetime to Newtonian mechanics formulated in Galilean (or Newton-Cartan) spacetime, as they did not have the relevant internal interpretation to hand (cf. [13, p. 854]). This resistance to *ab initio* declarations of physical equivalence is, of course, very consonant with the motivational approach—to which, for such reasons, we take Butterfield and De Haro to subscribe.

That said, the above, extended passage from Butterfield is ambiguous between a 'confident' version of motivationalism, according to which the internal interpretation on which dual (unextendable) theories may be regarded as being physically equivalent always exists, and a 'cautious' version of motivationalism, according to which there is no guarantee that such an internal interpretation exists. We take one of the central contributions of the present paper to be the delineating of these distinct sub-views, hitherto overlooked in the literature.[75]

[75] We thank Jeremy Butterfield and Sebastian De Haro for discussion on these matters; in private communica-

In our view, the cautious version of the motivational approach is to be preferred. Let us explain why; we have, in particular, three points to make. First, it seems to us possible to envisage (admittedly artificial) cases in which duality-related models do *not* possess sufficiently rich common structure to afford a coherent physical interpretation: imagine that the two models agree on empirical substructures, in addition to having extra structure, which is not isomorphic between the two models. In that case, the only 'common' structures may be the empirical substructures of both models—but the empirical substructures alone are insufficient for a *realist* understanding of what the mathematical model is supposed to represent in the world.

Second, some philosophers, most notably Dasgupta, would maintain that we do *not* yet have a coherent explication of the common ontology underlying models of Newtonian mechanics related by static shifts, or models of GR related by hole diffeomorphisms—for, Dasgupta argues, standard responses such as sophisticated substantivalism make appeal to problematic "bare modal claims" [12, p. 120]; moreover, any alternative approach, such as an appeal to so-called 'Einstein algebras' (cf. e.g. [25, ch. 9]), will face similar problems. In such a case, for authors such as Dasgupta (who, incidentally, also endorses something like the motivational approach to symmetries—cf. [13, pp. 853-854]), no such coherent explication of the common ontology underlying the models in question is forthcoming; nor is it guaranteed to exist. Thus, this particular case illustrates one of the ways in which one would be pushed towards the more cautious motivational view.

Third and finally, our preferred version of the motivational approach is, we contend, the most epistemically cautious interpretative attitude possible towards models of physical theories related by symmetries/dualities.[76] In our view, it is therefore unreasonable to demand that the burden of justification lies with this position; rather, there exists a positive burden of justification—here to argue that there always exists a coherent explication of the common ontology underlying symmetry- or duality-related models—for advocates of riskier approaches, such as the confident motivational approach, or the interpretational approach.

tion, both have indicated that they favour the cautious motivational approach.

[76] There is some analogy here with van Fraassen's constructive empiricism [71], often justified on the grounds that "belief in the empirical adequacy of accepted theories [is] the weakest attitude one can attribute to scientists at the same time that one is still able to make sense of their scientific activity" [47, §2.2].

6 Consensus

In §1, we highlighted two oft-advanced reasons for regarding duality-related models as being physically equivalent in the absence of a coherent explication of their common ontology: (1) such a position—the interpretational approach to dualities—fits with the parallel interpretational approach to symmetries; (2) such a position accords with a perceived consensus in contemporary theoretical physics. Up to this point, we have focussed on arguing against (1)—by arguing against the interpretational approaches to both symmetries and dualities, and for their motivational alternatives. In this section, we turn to (2), giving evidence that the interpretational approach to dualities does not, in fact, represent a consensus in the relevant areas of theoretical physics.

To begin, recall from §5.3 De Haro's observation that, in physics, the term 'duality' "is attached to the equivalence of the formal structures of the theories, regardless of their interpretations" [15, p. 12]. That is, in theoretical physics the term 'duality' is often applied to theories which are solely theoretically equivalent—physical equivalence notwithstanding. It is not hard to identify explicit evidence for De Haro's claim in the literature. For example, Vafa defines a duality thus: [73, pp. 4-5][77]

> Consider a physical system Q (which I will not attempt to define). And suppose this system depends on a number of parameters $[\lambda_i]$. Collectively we denote the space of the parameters λ_i by $\mathcal{M}$ which is usually called the moduli space of coupling constants of the theory. ... Typically physical systems have many observables which we could measure. Let us denote the observables by $\mathcal{O}_\alpha$. Then we would be interested in their correlation functions ... The totality of such observables and their correlation functions determine a physical system. Two physical systems $Q\left[\mathcal{M}, \mathcal{O}_\alpha\right]$, $\tilde{Q}\left[\tilde{\mathcal{M}}, \tilde{\mathcal{O}}_\alpha\right]$ are dual to one another if there is an isomorphism between $\mathcal{M}$ and $\tilde{\mathcal{M}}$ and $\mathcal{O} \leftrightarrow \tilde{\mathcal{O}}$ respecting all the correlation functions.[78]

Vafa's focus is clearly upon the formal, mathematical equivalence of the two theories in

[77] For philosophical discussions drawing upon Vafa's definition of a duality, see [41, §3] and [59, §2].

[78] Here, by a 'physical system' Q, we can understand Vafa to mean a physical theory $\mathcal{T}$; one may then identify $\mathcal{M}$ with our reduced space of DPMs, $\tilde{\mathcal{D}}$ (note that the tilde here refers to the quotienting of $\mathcal{D}$ by gauge-equivalent models, rather than to the space of DPMs of the dual theory). By $\mathcal{O}$, we understand Vafa to mean the set of observables $\mathcal{O}_\alpha$ for the theory in question.

question (i.e. upon formal equivalence, in the sense of §4.1). If one understands correlation functions as being the empirical substructures of the theories under consideration (in the sense of [71, pp. 67ff.]), then one may be able to argue in addition that Vafa is interested in the empirical equivalence of such theories. However, for our purposes the essential point here is that no claim regarding the physical equivalence of duality-related models is advanced. This accords with the view that, in the physics literature, the term 'duality' is often applied to cases of theoretical equivalence, without (explicit) assumptions being made regarding the physical equivalence of the models in question.

For a second piece of evidence to this end, consider the following quote from Maldacena [38, p. 61], made in the context of the AdS/CFT correspondence:

> What does it really mean for the two [dual] theories to be equivalent? First, for every entity in one theory, the other theory has a counterpart. The entities may be very different in how they are described by the theories: one entity in the interior might be a single particle of some type, corresponding on the boundary to a whole collection of particles of another type, considered as one entity. Second, the predictions for corresponding entities must be identical. Thus, if two particles have a 40 percent chance of colliding in the interior, the two corresponding collections of particles on the boundary should also have a 40 percent chance of colliding.

Maldacena is making two claims here. First, he is saying that, given two dual theories, every entity in each theory has a 'counterpart' entity in the corresponding dual theory. Second, he is saying that dual theories must be empirically equivalent. With regard to the first point: to say that each such entity has a 'counterpart' in the corresponding dual theory is, of course, very different from saying that they are truly one and the same thing. (The relation of counterparthood is not the identity relation!) And with regard to the second point: empirical equivalence is, of course, very different from physical equivalence. (Physical equivalence implies empirical equivalence, but not vice versa.)

To be clear, it is not our intention here to try to answer the (tricky) question of when exactly physicists do and do not call an isomorphism between spaces of solutions a duality. Rather, our purpose in this section is less ambitious: we seek merely to emphasise that physicists do not *invariably* regard dual models as being physically equivalent, and (relatedly) that they do not *invariably* regard the notion of 'duality' as involving the notion of physical equivalence. To do this, it is clearly not required that we get a grip on the question of when exactly physicists call an isomorphism between spaces of solutions a 'duality'—though we admit that this is indeed an interesting question worthy of further scrutiny.

7 Close

In this paper, we have argued that it is not invariably legitimate to regard duality-related models as being physically equivalent; rather, the existence of a duality at most *motivates one to seek* a coherent explication of the ontology underpinning their physical equivalence—and only once such an explication is secured may one indeed take the models in question to be physically equivalent. In order to achieve this goal, we have both (in §5) appealed to an analogous distinction in the case of dualities to that between the interpretational and motivational approaches to symmetry transformations; and argued (in §6)—for what it is worth—that physicists' understanding of duality-related models *vis-à-vis* their physical equivalence is less clear-cut than some philosophers take it to be. The moral of our investigations is a familiar one in the philosophy of physics: that the situation—here regarding dualities—is less straightforward than one might at first think. This paper may be judged a success to the extent that the (in our view) over-simplified, false impression that duality-related models may always be taken to represent the same physical state of affairs is dispelled.

Acknowledgements

We are indebted to Jeremy Butterfield (thrice over!), Sebastian De Haro, Neil Dewar, Niels Martens, Keizo Matsubara, Tushar Menon, Teruji Thomas, and several anonymous referees, for valuable comments on drafts of this paper, and are also grateful to Oliver Pooley and David Wallace for helpful discussions. J.R. is supported by an AHRC studentship at the University of Oxford, and is also grateful to Hertford College, Oxford for a graduate senior scholarship. T.M.-N. is supported by a stipendiary lectureship at Oriel College, Oxford.

References

[1] Thomas William Barrett and Hans Halvorson, "Glymour and Quine on Theoretical Equivalence", Journal of Philosophical Logic 45(5), pp. 467-483, 2016.

[2] Thomas William Barrett and Hans Halvorson, "Morita Equivalence", The Review of Symbolic Logic 9(3), pp. 556-582, 2016.

[3] David Baker, "Symmetry and the Metaphysics of Physics", Philosophy Compass 5, pp. 1157-1166, 2010.

[4] Katrin Becker, Melanie Becker, and John Schwarz, *String Theory and M-Theory: A Modern Introduction*, Cambridge: Cambridge University Press, 2007.

[5] Gordon Belot, "The Principle of Sufficient Reason", The Journal of Philosophy 98, pp. 55-74, 2001.

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

[7] Gordon Belot, "Fifty Million Elvis Fans Can't be Wrong", forthcoming in Noûs, 2017.

[8] Harvey R. Brown, *Physical Relativity: Space-Time Structure From a Dynamical Perspective*, Oxford: Oxford University Press, 2005.

[9] Jeremy Butterfield, "On Dualities and Equivalences Between Physical Theories", 2018.

[10] Elena Castellani, "Duality and 'Particle' Democracy", Studies in History and Philosophy of Modern Physics 59, pp. 100-108, 2017.

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

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

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

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

[15] Sebastian De Haro, “Spacetime and Physical Equivalence”, available at http://philsci-archive.pitt.edu/12279/, 2016.

[16] Sebastian De Haro, “The Heuristic Function of Duality”, Synthese, 2018. (Forthcoming.)

[17] Sebastian De Haro and Jeremy Butterfield, “A Schema for Dualities, Illustrated by Bosonization”, in J. Kouneiher (ed.), *Foundations of Mathematics and Physics One Century After Hilbert*, Berlin: Springer, 2018.

[18] Sebastian De Haro, Daniel R. Mayerson and Jeremy N. Butterfield, “Conceptual Aspects of Gauge/Gravity Duality”, Foundations of Physics 46(11), pp. 1381-1425, 2016.

[19] Sebastian De Haro, Nicholas Teh and Jeremy N. Butterfield, “Comparing Dualities and Gauge Symmetries”, Studies in History and Philosophy of Modern Physics 59, pp. 68-80, 2017.

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

[21] Neil Dewar, “Interpretation and Equivalence; or, Equivalence and Interpretation”, in E. Curiel and S. Lutz (eds.), *The Semantics of Theories*, 2017. (Forthcoming.)

[22] Neil Dewar, “Maxwell Gravitation”, Philosophy of Science 85, pp. 249-270, 2018.

[23] Paul Dirac, *The Principles of Quantum Mechanics*. Oxford: Oxford University Press, 1930.

[24] Anthony Duncan, *The Conceptual Framework of Quantum Field Theory*, Oxford: Oxford University Press, 2017.

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

[26] John Earman and John Norton, “What Price Spacetime Substantivalism? The Hole Story”, British Journal for the Philosophy of Science 38(4), pp. 515-525, 1987.

[27] Michael Esfeld and Vincent Lam, “Moderate Structural Realism About Spacetime”, Synthese 160(1), pp. 27-46, 2011.

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

[29] Hilary Greaves and David Wallace, "Empirical Consequences of Symmetries", British Journal for the Philosophy of Science 65, pp. 59-89, 2014.

[30] Richard Healey, *Gauging What's Real: The Conceptual Foundations of Contemporary Gauge Theories*, Oxford: Oxford University Press, 2009.

[31] Nick Huggett, *Space from Zeno to Einstein*, Cambridge, MA: MIT Press, 1999.

[32] Nick Huggett, "Target Space $\neq$ Space", Studies in the History and Philosophy of Modern Physics 59, pp. 81-88, 2017.

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

[34] Eleanor Knox, "Newtonian Spacetime Structure In Light of the Equivalence Principle", British Journal for the Philosophy of Science 65(4), pp. 863-880, 2014.

[35] James Ladyman, "On the Identity and Diversity of Objects in a Structure", Proceedings of the Aristotelian Society 81, pp. 23-43, 2007.

[36] David Malament, *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*, Chicago: University of Chicago Press, 2012.

[37] Juan Maldacena, "The Large N Limit of Superconformal Field Theories and Supergravity", Advances in Theoretical and Mathematical Physics 2, pp. 231-252, 1998.

[38] Juan Maldacena, "The Illusion of Gravity", Scientific American, November 2005.

[39] Keizo Matsubara, "Realism, Underdetermination and String Theory Dualities", Synthese, pp. 471-489, 2013.

[40] Keizo Matsubara and Lars-Göran Johansson, "Spacetime in String Theory: A Conceptual Clarification", 2016.

[41] Keizo Matsubara and Chris Smeenk, "Dualities and Effective Ontology", 2016.

[42] Tim Maudlin, "Buckets of Water and Waves of Space", Philosophy of Science 60(2), pp. 183-203, 1993.

[43] Tim Maudlin, *Philosophy of Physics: Space and Time*. Princeton, NJ: Princeton University Press, 2012.

[44] Charles Misner, Kip Thorne, and John Wheeler, *Gravitation*, San Francisco: Freeman & Co., 1973.

[45] Thomas Møller-Nielsen, "Invariance, Interpretation, and Motivation", Philosophy of Science 84, pp. 1253-1264, 2017.

[46] Thomas Møller-Nielsen and James Read, "Redundant Epistemic Symmetries", 2017.

[47] Bradley Monton and Chad Mohler, "Constructive Empiricism", in *The Stanford Encyclopedia of Philosophy*, 2017.

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

[49] Michael E. Peskin and Daniel V. Schroeder, *An Introduction to Quantum Field Theory*, Reading, MA: Perseus Books, 1995.

[50] Joseph Polchinski, "Dualities of Fields and Strings", Studies in History and Philosophy of Modern Physics 59, pp. 6-20, 2017.

[51] Oliver Pooley, *The Reality of Spacetime*, D.Phil. Thesis, University of Oxford, 2002.

[52] Oliver Pooley, "Points, Particles, and Structural Realism", In D. Rickles, S. French & J. Saatsi (eds.), *The Structural Foundations of Quantum Gravity.* Oxford: Oxford University Press, 2006.

[53] Oliver Pooley, "Substantivalist and Relationist Approaches to Spacetime", in R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics*, Oxford: Oxford University Press, 2013.

[54] Oliver Pooley, "Background Independence, Diffeomorphism Invariance, and the Meaning of Coordinates", in D. Lehmkuhl, G. Schiemann and E. Scholz (eds.), *Towards a Theory of Spacetime Theories*, Birkhäuser, 2017.

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

[56] Willard Van Orman Quine, “On the Reasons for Indeterminacy of Translation”, Journal of Philosophy 67(6), pp. 178-83, 1970.

[57] Willard Van Orman Quine, “On Empirically Equivalent Systems of the World”, Erkenntnis 9, pp. 313-328, 1975.

[58] James Read, “Background Independence in Classical and Quantum Gravity”, B.Phil. Thesis, University of Oxford, 2016.

[59] James Read, “The Interpretation of String-Theoretic Dualities”, Foundations of Physics 46(2), pp. 209-235, 2016.

[60] Dean Rickles, “A Philosopher Looks at String Dualities”, Studies in History and Philosophy of Modern Physics 42, pp. 54-67, 2011.

[61] Dean Rickles, “AdS/CFT Duality and the Emergence of Spacetime”, Studies in History and Philosophy of Modern Physics 44, pp. 312-320, 2013.

[62] Dean Rickles, “Mirror Symmetry and Other Miracles in Superstring Theory”, Foundations of Physics 43, pp. 54-80, 2013.

[63] Dean Rickles, “Dual Theories: ‘Same but Different’ or ’Different but Same’?”, Studies in History and Philosophy of Modern Physics, pp. 62-67, 2017.

[64] Simon Saunders, “Indiscernibles, General Covariance and Other Symmetries: The Case for Non-Reductive Relationalism”, in A. Ashketar *et al.* (eds.), *Revisiting the Foundations of Relativistic Physics: Festschrift in Honour of John Stachel*, Dordrecht: Kluwer Press, 2003.

[65] Simon Saunders, “Rethinking Newton’s *Principia*.” Philosopy of Science 80(1), pp. 22-48, 2013.

[66] Mark Srednicki, *Quantum Field Theory*, Cambridge: Cambridge University Press, 2007.

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

[68] Nicholas Teh, “Galileo’s Gauge: Understanding the Empirical Significance of Gauge Symmetry”, Philosophy of Science 83(1), pp. 93-118, 2016.

[69] Nicholas Teh, "Recovering Recovery: On the Relationship Between Gauge Symmetry and Trautman Recovery", Philosophy of Science 85, pp. 201-224, 2018.

[70] Andrzej Trautman, "Foundations and Current Problems of General Relativity", in S. Deser & K. W. Ford (eds.), *Lectures on General Relativity*, New Jersey: Englewood Cliffs, Prentice-Hall, 1965.

[71] Bas C. Van Fraassen, *The Scientific Image*, Oxford: Oxford University Press, 1980.

[72] Bas C. Van Fraassen, *Laws and Symmetry*, Oxford: Oxford University Press, 1989.

[73] Cumrun Vafa, "Geometric Physics", in *Proceedings of the International Congress of Mathematics, Vol. 1*, G. Fischer and U. Rehmann (eds.), 1998.

[74] David Wallace, "Fields As Bodies: A Unified Presentation of Spacetime and Internal Gauge Symmetry", 2015. arXiv:1502.06539.

[75] David Wallace, "Fundamental and Emergent Geometry in Newtonian Physics", British Journal for the Philosophy of Science, 2017. (Forthcoming.)

[76] James Owen Weatherall, "Are Newtonian Gravitation and Geometrized Newtonian Gravitation Theoretically Equivalent?", Erkenntnis 81(5), pp. 1073-1091, 2016.

[77] James Owen Weatherall, "Maxwell-Huygens, Newton-Cartan, and Saunders-Knox Spacetimes", Philosophy of Science 83(1), pp. 82-92, 2016.

[78] James Owen Weatherall, "Understanding Gauge", Philosophy of Science 85(5), pp. 1039-1049, 2016.

[79] James Owen Weatherall, "Categories and the Foundations of Classical Field Theories", in E. Landry (ed.), *Categories for the Working Philosopher*, Oxford: Oxford University Press. (Forthcoming.)

[80] Hermann Weyl, *Symmetry*, Princeton, NJ: Princeton University Press, 1952.

[81] Edward Witten, "String Theory Dynamics in Various Dimensions", Nuclear Physics B 443, pp. 85-126,1995.