

The Role of Symmetry in the Interpretation of Physical Theories

Adam Caulton
University of Cambridge
adam.caulton@gmail.com
+44 7876 425077

Contents

1	Introduction	2
2	States, quantities and symmetries: the first triangle	3
2.1	What is a physical theory?	3
2.2	Symmetries	7
2.3	Analytic and synthetic symmetries	8
3	A model of theory interpretation: the second triangle	14
3.1	The need for interpretative constraints	14
3.2	A proposal	16
4	Conclusion	19

1. Introduction

Physical theories—if we take them literally—tell us what world is like, or at least what it might have been like. For example, the consensus is that special relativity, if taken as a literal description of the world, tells us that the world is not a sequence of three-dimensional instants, objectively unfolding over time, but rather a four-dimensional continuum, in which there is no objective temporal order between spacelike separated events. Meanwhile, there seems to be little consensus about what quantum mechanics, taken literally, tells us about the world. Clearly, we cannot extract an account of the physical world from a theory unless we first understand it. But how do we come to understand a theory—that is, how do we *interpret* a theory—if (as we suspect) it describes things that we have never before even *thought* of? In other words: How are we supposed to know what a theory is aiming at—what it represents as holding true—if the candidates for its target can only be understood by the descriptions given of them *by that very theory*?

Van Fraassen (2008) has recently argued that this predicament prevents us from giving any literal interpretation to a physical theory, except insofar as it describes observable entities. In this paper, I will marshal some considerations from philosophy of language and philosophy of science to argue that a literal interpretation for a physical theory *can* be found—one that describes observables and non-observables alike—under the guidance of the right interpretative constraints. These constraints are Leibnizian in letter, for they recommend the elimination of distinctions without a difference, in much the same way that Leibniz did in his critique of Newtonian absolute space in his correspondence with Clarke (Alexander 1998). They appeal to *symmetries* of the theory being interpreted, as guides for what to eliminate. (As I will explain, symmetries are transformations on a theory’s space of states or solutions that preserve certain quantities defined over that space.) But the constraints are not Leibnizian in spirit, since the motivation for them is semantic rather than metaphysical or theological—in particular, I will not appeal to the Principle of Sufficient Reason. In fact the spirit of this paper owes at least as much to Reichenbach (1958, 1965) and Carnap (1966) as to Leibniz.

The paper is divided into two main parts, each focussed on what I will call ‘a triangle of concepts’. In the first part (Section 2), I give a general account of physical theories and their symmetries. There I introduce my first triangle of concepts: these are a theory’s states, quantities and symmetries. I propose a dichotomy between symmetries into what I call *analytic* and *synthetic*, so-called because of analogies with the familiar characterisation of propositions in those terms. This dichotomy, applied to symmetries, is crucial for my recommendations in part two for interpreting a physical theory.

The second part (Section 3) contains a discussion of theory interpretation, and makes proposals for how to interpret a theory that is taken to be about (amongst other things) as-yet-unconceived entities. There I introduce my second triangle of concepts: a theory’s formalism, its subject matter, and the representation relation that exists between them. The second triangle is linked to the first: for which symmetries count as analytic and which count as synthetic will hang on the details of this representation relation, and *vice versa*. Therefore a constraint on the symmetries will serve as a constraint on the representation relation, hopefully one strong enough to determine a unique interpretation for the theory.

Here I must immediately qualify the sense of ‘interpretation’ under which it would be plausible to claim that a unique one might be found. As we shall see in Section 2, my main concern is the representation relation that may hold between possible states, or histories, and physical quantities (i.e. properties and relations) on the one hand, and their mathematical representatives on the other: in particular whether it is one-to-one or one-to-many. In these terms, the sense of interpretation I have in mind must be rather coarse-grained: it is intensional, as opposed to hyper-intensional. As such it will leave many metaphysical questions unanswered—specifically ones which make no difference to the variety of possible states or histories. For example, an interpretation will not settle whether it is better to interpret unit electric charge as a universal or a collection of intrinsically similar tropes, or whether gravitational attraction in Newtonian particle mechanics is a direct relation between particles or mediated by a gravitational field.¹ However, an interpretation in my sense *will* settle questions of ontology insofar as these lead to differences in the corresponding variety of possibilities. These include the existence or otherwise of: absolute location, orientation or velocity; ‘haecceities’ for spacetime points or elementary particles; and the electromagnetic four-potential. The main purpose of the paper is to argue that, by maximising the number of the theory’s analytic symmetries, subject to empirical adequacy, a unique interpretation—in this coarse-grained sense of ‘interpretation’—can be found.

2. States, quantities and symmetries: the first triangle

2.1. What is a physical theory?

I should begin by saying a few words about what I take a physical theory to be. I want to be general enough to incorporate a variety of theories, both classical and quantum. My account will not be particularly novel, nor even adequate for all purposes; but it will be adequate for the purposes of this paper. I will use some ideas and techniques associated with the syntactic view of theories

¹I am grateful to an anonymous referee for this example.

(e.g. I partly characterise a theory by its axioms) and some associated with the semantic view (e.g. I partly characterise a theory by its models): I see no reason to opt for one approach over the other.²

So I propose to take a physical theory to comprise the following four components:

1. A *language* \mathcal{L} . This is intended to be the language in which the equations of motion of the theory are expressed; but it should also incorporate the fragment of simple English that surrounds these equations. (E.g. ‘Let \mathcal{M} be a differentiable manifold,’ or ‘The seventh factor Hilbert space represents the possible states for particle seven.’) I propose to give a semantics for this language that follows (at least in spirit) the general scheme laid down by Montague (1970) and Lewis (1970b). That is, it is *intensional* and *compositional*.

Briefly, intensionality means that linguistic items—particularly sentences—are assigned intensions. For many linguistic items (specifically, the ones which may be assigned extensions—i.e., sentences and noun-phrases, whose extensions are truth-values and worldly objects, respectively), the intension is a function from *indices* to extensions. In Montague’s theory, an index is a possible world, construed as a Tarskian model, or an ordered sequence comprising a possible world and appropriate context-determining parameters, such as the time and place of utterance.³ However, for the purposes of interpreting a physical theory, I propose that we here stray from Montague by supplanting Tarskian models with the possible states or histories already belonging to the mathematical formalism of our given theory of interest; see below.

Compositionality is achieved by breaking down whole sentences (e.g. ‘the electron is spin-up’) into sub-sentential components whose intensions have a function/argument structure. So, for example, the sentence ‘the electron is spin-up’ (whose intension maps any possible state in which the electron is spin-up to *true* and all others to *false*) is broken down into the noun-phrase ‘the electron’ (whose intension maps any possible state to the electron, if there is a unique one, in that state) and the verb-phrase ‘is spin-up’, whose intension is a function from noun-phrase-intensions to sentence-intensions; specifically, it maps the intension of ‘the electron’ to the intension of ‘the electron is spin-up’.

An important collection of noun-phrases used in physical theories are those used to refer to determinates, such as ‘2kg in mass’, ‘distance of 3km’, ‘spin-up’, etc. This suggests that we include amongst our worldly objects not just things like particles or spacetime points, but also masses, distances and vector of spin. These *dimensioned values* lie in some “logical space”—à la van Fraassen (1967) and Stalnaker (1979)—which has a mathematical parametrisation, e.g. by the real line or the vector space of anti-symmetric rank-2 tensors. Of course, any gloss given to these determinates at this stage is at best provisional, since we have not yet interpreted the theory! I return to determinates below, in the discussion of quantities.

2. A space of *mathematical states* \mathcal{S} . Almost all theories are equipped with a space of its own “possible worlds”.⁴ They have many commonalities with Tarskian models and Lewis’s (1986) possible worlds, most crucial among them being that their role is to represent possibilities

²The classic account of the failings of the syntactic view is Suppe (1974). For details of the semantic view, see e.g. French & Ladyman (1999), van Fraassen (2000) and Frigg & Hartmann (2008). For a compelling defence of the syntactic view, see Lutz (2010).

³For a comprehensive discussion of the treatment of context gestured at here, see Lewis (1980)

⁴A notable exception may be quantum field theories—at least on the algebraic approach—each of whose possible states cannot be captured by any single separable Hilbert space.

regarding the actual (physical) world: i.e., ways the world and its constituents might have been. However, there are two crucial differences between these states and both Tarskian models and Lewisian possible worlds.

The first crucial difference is that, while the means by which Tarskian models and Lewisian possible worlds represent possibilities is (at least very often taken to be) well-defined and unproblematic, the representation relation between the mathematical states of a physical theory’s formalism and corresponding concrete possibilities is far from obvious. Indeed, a major task in the interpretation of any physical theory is precisely settling on a unique representation relation; I return to this below in section 3.1. As a result, the interpretation of a given physical theory will not be complete when a Montague semantics is found for its language: for the intensions in any such semantics will make ineliminable reference to mathematical states whose physical correlates have not yet been laid down. It is only when the further task of settling the representation relation between mathematical states and physical possibilities has been completed that we can say in all good conscience that we have provided a semantics for our theory.

The second crucial difference is that in many theories—in particular classical and quantum particle mechanics—the job of mathematical states is in representing *instantaneous* time-slices of the physical world, rather than temporally extended histories. Histories are often represented in these theories by *trajectories*: i.e. maps from (some segment of) \mathbb{R} (or, better: a temporal “logical space”, perhaps taken to be isomorphic to \mathbb{R}), into \mathcal{S} . However, in other theories, particularly ones without a single objective time parameter (e.g. general relativity), the job of the mathematical states *is* to represent entire histories.⁵

The elements of \mathcal{S} , or the trajectories on \mathcal{S} , represent possible states or worlds under the widest conception of possibility discussed here, namely *kinematic possibility*. We expect the elements or trajectories on \mathcal{S} to be whittled down by equations of motion and initial or boundary conditions. But I assume here that a satisfactory interpretation of a physical theory requires settling the representation relation for all kinematically possible states.

The approaches to laying down a representation relation between the mathematical states of \mathcal{S} and concrete possibilities fall into two broad types, which I call *indirect* and *primitive*. On the indirect type of approach, we aim to specify a representation relation by associating the states (or trajectories) of \mathcal{S} with *other* formal representatives of possibilities, whose relationship to those possibilities is taken to be unproblematic. For example, we may seek to associate with every element of \mathcal{S} a Tarskian model; the theory’s ontological commitments could then simply be read off of the contents of the domain, as envisaged by Quine (1948).

But there is also the primitive type of approach, in which the representation relation is settled directly. Here is the source of a major problem: how do we specify a relation between two types of thing when, not only are we not knowingly acquainted with one of the types of thing in question, but our understanding of those things is provided *solely* by the first type of things? This is an issue discussed in section 3.1.

3. A set of *quantities* \mathcal{Q} . Quantities are functions from \mathcal{S} —or the space of trajectories through \mathcal{S} —into some value-space \mathcal{V} , which is a “logical space” in the van-Fraassen-Stalnaker sense mentioned above. It is likely that a sensible choice for \mathcal{Q} will be more restricted than the

⁵Attempts to represent quantum *histories* as vector states in a Hilbert space, rather than as maps from \mathbb{R} into a Hilbert space, have been made by e.g. Johnson (1969) and Broyles (1970).

set of all functions from \mathcal{S} into \mathcal{V} and that \mathcal{Q} may be endowed with a structure beyond its merely being a set. (For example, the quantities in the Hamiltonian formulation of classical particle mechanics are commonly taken to be *smooth* real-valued functions on the classical phase space. The set of such quantities not only form an infinite-dimensional vector space through pointwise addition; they also form a Poisson algebra through pointwise multiplication and the Poisson bracket.)

Now let us anticipate the interpretative project somewhat. Some subset \mathcal{Q}_Φ of \mathcal{Q} will be privileged in virtue of the fact that its members register physical differences; i.e. \mathcal{Q}_Φ supervenes on the physical properties or relations (which have not yet been mentioned).⁶ Let us call the members of \mathcal{Q}_Φ the *physical quantities* of the theory, even though, strictly speaking, they are mathematical functions.

There is much we can say already about what we should expect \mathcal{Q}_Φ to contain. For most theories, we might expect it to contain the theory’s Hamiltonian function, H , since it governs the theory’s dynamics. In fact this consideration is not compelling, since H may be subject to scale transformations which might be thought to preserve the physics.⁷ However, we can say for certain that in the Lagrangian formulation of classical particle mechanics \mathcal{Q}_Φ will contain the Euler-Lagrange functions $E_\mu := \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^\mu} - \frac{\partial L}{\partial q^\mu}$. This is because solutions are characterised by the equations $E_\mu = 0$, and a difference in this value will make the difference between obeying the dynamical equations and violating them: surely a physical difference if anything is.

4. A set of *axioms* T . These are the equations of the theory, written in \mathcal{L} and expressing dynamical laws and initial and/or boundary conditions. Their purpose is to pick out privileged models constructed from the states in \mathcal{S} . These privileged models may correspond to trajectories through \mathcal{S} , or just states of \mathcal{S} . We may call them the *dynamically possible worlds*; they comprise a proper subset of the kinematically possible worlds.

The axioms in T are put to use in the generation of predictions, i.e. the derivations of values, or constraints on the values, of the quantities in \mathcal{Q} . For example, in general relativity, a specification of the “matter field” T_{ab} for all of history, and of the “metric field” g_{ab} on the asymptotic boundary, combined with the the Einstein field equations, will generate a class of predictions for the value of g_{ab} on all spacetime—although, famously, no *unique* specification will typically be forthcoming.⁸

That concludes my brief overview of what we may take physical theories to be, for the purposes

⁶The notion of supervenience I am using here is a relation between sets of properties: \mathcal{A} supervenes on \mathcal{B} if and only if there is no change in (which objects instantiate) \mathcal{A} without there being a corresponding change in (which objects instantiate) \mathcal{B} . Note that this does not require *every* quantity in \mathcal{Q}_Φ to register *every* physical difference; rather it requires that if there be a change in *some* quantity in \mathcal{Q}_Φ , then it correspond to a change in *some* physical property or relation. I do not demand that the physical properties and relations supervene on \mathcal{Q}_Φ as well as *vice versa*, since it may not be the purpose of the theory to represent all of the physical information.

⁷An example of this is the simple harmonic oscillator in classical mechanics. A scale transformation on the displacement co-ordinate $q \mapsto kq$ induces the transformation $H \mapsto k^2H$, but if the transformation is interpreted as a change of length unit (from centimetres to inches, for example), then this clearly makes no physical difference.

⁸This is Einstein’s famous ‘hole argument’. For just a sample of the extensive literature on this, see e.g. Earman & Norton (1987), Butterfield (1989) and Rynasiewicz (1994). But also see Weatherall (2015) for an argument that, on the appropriate notion of ‘equivalence’, any two predictions are in fact equivalent, and so the prediction is essentially unique.

of this paper. Let us now turn to symmetries.

2.2. Symmetries

Generally speaking, a symmetry is a transformation on an object that preserves some salient feature of that object. The symmetry of a snowflake consists in the fact that there are six imaginary lines passing through the snowflake’s centre, reflections through which (approximately) preserve the snowflake’s shape. In logic, a relation defined on a set of objects is called symmetric if, for any pair, it holds both ways or not at all: that is, its holding or not is preserved under interchange of the argument places.

In this paper, I am concerned almost exclusively with the symmetries of one object: the space of mathematical states \mathcal{S} (or the corresponding space of trajectories). Thus I will define a symmetry as a transformation—a bijection—on \mathcal{S} which preserves certain salient features of \mathcal{S} . The salient features are the values of at least some of the quantities in \mathcal{Q}_Φ , i.e. the physical quantities. Thus a symmetry is a bijection on the space of mathematical states which preserves the values of at least some of the physical quantities (from now on I will just say, ‘preserves the physical quantities’). Let \mathcal{Q}' be the set of physical quantities that we are interested in preserving, and let \mathcal{A}' be the set of symmetries which preserve \mathcal{Q}' . Then

$$\text{For all } a, a \text{ is in } \mathcal{A}' \text{ iff } (\forall Q \in \mathcal{Q}')(\forall s \in \mathcal{S}) Q(a(s)) = Q(s), \quad (1)$$

so:

$$(\forall a \in \mathcal{A}')(\forall Q \in \mathcal{Q}')(\forall s \in \mathcal{S}) Q(a(s)) = Q(s). \quad (2)$$

This equation acts as a constraint for the triangle of states, quantities and symmetries.

As we would expect, the symmetries in \mathcal{A}' form a group.⁹ If we prefer, we may consider a theory’s symmetries as a transformation on its *quantities*, rather than its *states*. For every Q in \mathcal{Q} , we define its “pullback” a^*Q under a with the condition: $(\forall s \in \mathcal{S}) (a^*Q)(s) = Q(a(s))$. Then we may rephrase the triangulation constraint (2) as

$$(\forall a \in \mathcal{A}')(\forall Q \in \mathcal{Q}') a^*Q \equiv Q, \quad (3)$$

where ‘ \equiv ’ expresses identity of functions, rather than (mere) identity of their values in some state.

It is clear from (2) and (3) how the space of states, the set of relevant quantities, and the group of symmetries depend on one another. I will now quickly run through two salient sorts of negotiations that we might undergo with this triangle.

- Being more strict about which quantities go in \mathcal{Q}' leads to more state space bijections falling in \mathcal{A}' . Conversely, the more we put into \mathcal{Q}' , the fewer transformations we find in \mathcal{A}' . (Similarly,

⁹*Proof:* (i) The identity transformation, defined by $id(s) := s$ for all $s \in \mathcal{S}$, is in \mathcal{A}' , since the right-hand side of the bi-conditional in (1) holds trivially for it. (ii) If a is in \mathcal{A}' , then so is a^{-1} , since $(\forall Q \in \mathcal{Q}')(\forall s \in \mathcal{S}) Q(a(s)) = Q(s)$ iff $(\forall Q \in \mathcal{Q}')(\forall s \in \mathcal{S}) Q(s) = Q(a^{-1}(s))$, so long as \mathcal{S} is closed under a , which, since a is a bijection on \mathcal{S} , it is. (iii) If a_1 and a_2 are in \mathcal{A}' , then so is their composition $a_2 \circ a_1$, which can be seen as follows. $(\forall Q \in \mathcal{Q}')(\forall s \in \mathcal{S}) Q(a_2(s)) = Q(s)$, since a_2 is in \mathcal{A}' . But, since \mathcal{S} is closed under a_1 , this is equivalent to $(\forall Q \in \mathcal{Q}')(\forall s \in \mathcal{S}) Q(a_2(a_1(s))) = Q(a_1(s))$. We also have that $(\forall Q \in \mathcal{Q}')(\forall s \in \mathcal{S}) Q(a_1(s)) = Q(s)$. These last two results jointly imply $(\forall Q \in \mathcal{Q}')(\forall s \in \mathcal{S}) Q((a_2 \circ a_1)(s)) = Q(s)$. *QED.*

we may control \mathcal{Q}' by driving \mathcal{A}' .) There are two extreme, trivialising cases. (1) The first occurs when we set $\mathcal{Q}' = \mathcal{Q}$, the full set of the theory's quantities. In this case $\mathcal{A}' = \{id\}$, since we would be requiring of a symmetry that it preserve *everything*; but preserving everything means changing nothing. (2) The second occurs when we set $\mathcal{Q}' = \emptyset$. In this case \mathcal{A}' is the full set of bijections on \mathcal{S} .

The interesting cases lie somewhere between these two extremes. One salient place to rest is to set $\mathcal{Q}' = \mathcal{Q}_\Phi$, the set of physical quantities—i.e. the quantities that, on the basis of their representing the physical properties and relations, register physical differences. More on this below.

- We may try to satisfy (2) by restricting \mathcal{S} , rather than \mathcal{Q}' or \mathcal{A}' . More specifically, we may jettison all the mathematical states except those for which $(\forall a \in \mathcal{A}') a(s) = s$; i.e. we can require that the mathematical states themselves be invariant under the relevant group of symmetries. Then (2) is satisfied, no matter how large the set of quantities \mathcal{Q}' . However, this dramatic reduction in \mathcal{S} leads anyway to dramatic consequences for \mathcal{Q}' . This is because the set \mathcal{Q} of quantities was originally defined on the domain \mathcal{S} ; and if we greatly reduce that domain, then we thereby induce a move to a (strictly speaking) different, but smaller, set of quantities.

In classical mechanics it would be disastrous to make this move, since the surviving mathematical states will themselves be highly symmetrical, and will likely not have sufficient wherewithal to represent an asymmetric physical world such as our own. However, in quantum mechanics, the move is far from disastrous. First: the superposition of states allows highly symmetric states to contain a great deal of information. In fact the Symmetrization Postulate, which demands that quantum states be invariant under permutations of factor Hilbert space labels, is just such an example of this move. Second: such highly symmetric states, corresponding to either bosons or fermions, actually suffice to represent all particles so far observed in the universe.¹⁰

2.3. Analytic and synthetic symmetries

Let us concentrate on the interesting case mentioned above, in which we set \mathcal{Q}' , the set of quantities we wish to preserve under transformations, to \mathcal{Q}_Φ , the set of physical quantities. Let the group of symmetries that preserve \mathcal{Q}_Φ be \mathcal{A}_Φ . Then, from (3),

$$(\forall a \in \mathcal{A}_\Phi)(\forall Q \in \mathcal{Q}_\Phi) a^*Q \equiv Q. \quad (4)$$

The symmetries in \mathcal{A}_Φ by definition effect no physical change, since there can be no change in \mathcal{Q}_Φ without a corresponding change in the physical properties and relations. Therefore, for each mathematical state s in \mathcal{S} , the equivalence class $\{a(s) \mid a \in \mathcal{A}_\Phi\}$ —i.e. the *orbit* of s , or the *group action*, under \mathcal{A}_Φ , on s —must represent only a *single* physical state. (To see that the orbits are equivalence classes, consider that ‘is related to by the action of an element of the group’ is an equivalence relation, in virtue of the characteristic properties of groups.) \mathcal{A}_Φ therefore induces a partition of \mathcal{S} in which each equivalence class represents just one physical state.

¹⁰For more details, see French & Redhead (1988) and Caulton & Butterfield (2012).

In quantum mechanics, we can go even further than this.¹¹ There, the symmetries in \mathcal{A}_Φ are unitary representations of a group. (Beware of the use of the term ‘representation’ here—it has nothing to do with the representation relation I discuss below!) Using a central result known as *Schur’s lemma*, it can be proven that the Hilbert space (our particular instantiation of \mathcal{S}) decomposes into orthogonal *subspaces* (i.e. not just equivalence classes of vectors, though each subspace will be the *span* of an orbit of vectors under \mathcal{A}_Φ). These subspaces are then the candidate representatives for single physical states.¹²

The existence of non-trivial symmetries in \mathcal{A}_Φ prompts a *quotienting* of our original space of mathematical states \mathcal{S} . The *quotient space* $\mathcal{S}/\mathcal{A}_\Phi$ is the space of the orbits (or subspaces) under \mathcal{A}_Φ . We may then define new quantities for the mathematical states in $\mathcal{S}/\mathcal{A}_\Phi$, by applying a rule of supervaluation. Any quantity defined on \mathcal{S} that is preserved under \mathcal{A}_Φ defines a unique, single-valued quantity defined on $\mathcal{S}/\mathcal{A}_\Phi$; all other quantities may be dropped. In this way, quotienting sheds redundant theoretical structure. The advantage of the quotient space is that it has itself no non-trivial symmetries which preserve all the physical quantities \mathcal{Q}_Φ . Thus it provides a more illuminating picture of what the space of *physical* states looks like, according to the theory.¹³

I propose to call the symmetries in \mathcal{A}_Φ the *analytic* symmetries. This is because of two significant and related parallels with analytic propositions (both must hold for the symmetry to be analytic). The first parallel is that analytic symmetries are physically empty, in that they do not produce a different physical state. This is due to their preserving *all* of the physical quantities. This is analogous to the characterisation of analytic propositions as not conveying any significant information (Carnap 1966).

The second parallel is that, for any transformation a , we can know whether or not a is an analytic symmetry solely in virtue knowing which quantities are in \mathcal{Q}_Φ , which is to say (as we shall see below): solely in virtue of knowing the representation relation that holds between the theory’s formalism and the physical properties and relations. (Note that this is already close to saying that a transformation’s being an analytic symmetry holds solely “in virtue of meaning”, as analytic propositions are commonly said to do.) We need to know no particular matters of fact to know whether or not a quantity belongs to \mathcal{Q}_Φ : in particular we do not need to know whether or not the theory’s axioms T are true. Equation (4) above is true in all kinematically possible worlds, not just the dynamically possible worlds. (Admittedly, this is a slightly strange thing to say, since equation (4) is a claim *about* all of the kinematically possible worlds. But it can be easily transformed into a claim to be assessed *at* each world, by asserting identity between the two quantities’ *values* instead of between the two quantities themselves. Identity of the values of functions for all arguments *is* identity of the functions.) This is analogous to the characterisation of analytic propositions as being true in all possible worlds (Carnap (1956); Lewis (1970b)).¹⁴

My claim that there is a parallel with analytic propositions here is weakened by the fact that

¹¹The following is a summary of results discussed in more detail in e.g. Tung (1985, Ch. 3).

¹²An argument to this effect in the case where $\mathcal{A}_\Phi \cong S_n$, the group of permutations on n objects, is given in Caulton & Butterfield (2012).

¹³For a philosophical discussion of quotienting, see Belot (2001). For more technical presentations, see Belot (2003) and Butterfield (2006).

¹⁴It is worthy of investigation whether equation (4) is logically entailed by the theory’s Carnap sentence, traditionally taken to embody the theory’s definitional content. I do not such an investigation here.

we do not actually assess whether (4) holds in all *logically* possible worlds. Rather, we restrict ourselves to the kinematically possible worlds, circumscribed by the theory. To claim that this is no real restriction, we would have to claim that the meanings of our terms render empty the idea of a kinematically impossible world. But this is patently not the case: the kinematically possible world of one theory are typically kinematically impossible by the light of another theory. I said that to know whether a symmetry was analytic, we need only know the representation relation between the theory and the physical properties and relations. That is true, but the setting up of the representation relation makes non-trivial assumptions, tantamount to assuming that the physical world is represented by at least one of the kinematically possible worlds. Therefore my use of the term ‘analytic’ must be taken with a pinch of salt. However, this does not detract from my main point that the analytic symmetries are importantly different from the rest. Even if the analytic symmetries do not hold as a matter of logical truth, still there is an important and clear sense in which they are fundamental to the theoretical framework: a change in the analytic symmetries means a change in what it is we may say the theory is representing. And this has rather more than a passing resemblance to a change in *language*.

The discrepancy between logically possible and kinematically possible worlds just mentioned echoes an observation made by Brading & Brown (2003, p. 99) regarding gauge symmetries in classical Lagrangian field theory:

‘While it is true that global gauge symmetry does not have the same direct empirical significance that global spacetime symmetries have, this does not imply that global gauge symmetry is without empirical content. *The very fact that a global gauge transformation does not lead to empirically distinct predictions is itself non-trivial.* In other words, the freedom in our descriptions is no “mere” mathematical freedom—it is a consequence of a physically significant structural feature of the theory.’

On the view defended here, the empirical content of the existence of global gauge symmetries amounts to a logically contingent claim about the physical ontology, which we take the theory to be making. Specifically, the claim is that certain elements in the mathematical formalism—in this case, one’s choice of co-ordinate system—have no physical correlate, such as haecceities for spacetime points. However, I disagree that such a claim is a ‘consequence of a physically significant structural feature of the theory’, if ‘structural’ here means formal, i.e. independent of a choice of representation relation. The claim is non-trivial precisely because the interpretation of the formalism is not compulsory.¹⁵

One final comment on the use of ‘analytic’: In fact it is common in physics and philosophy of physics circles to talk of “gauge” symmetries in much the same sense as that explicated for my proposed analytic symmetries; i.e. as transformations which preserve the physical state (e.g. Henneaux & Teitelboim 1992, pp. 3, 16-17). But ‘gauge symmetry’ is ambiguous, since gauge symmetries are also often defined purely formally, as a space-, time- or spacetime-dependent transformations in some internal space (expressed in the language of fibre bundles as a principle automorphism; see e.g. Isham 1999, p. 258). Therefore I propose to retain the term ‘gauge’ for the cluster of closely related formal notions, and adopt ‘analytic’ for the interpretative notion. It may indeed be claimed

¹⁵I am grateful to an anonymous referee for raising this issue.

that all gauge symmetries are analytic and all analytic symmetries are gauge. But establishing that claim would be a significant philosophical achievement: one that could be clear only in a language that refers to them with different names. Besides, it seems to me that there are both analytic symmetries that are not gauge and gauge symmetries that are not analytic. Permutations in many-particle quantum mechanics are surely an example of the former (see Caulton & Butterfield 2012); the revaluation of a currency in a foreign-exchange market may be an (admittedly *recherché!*) example of the latter.¹⁶

So much for analytic symmetries. Let us now move on to symmetries that are not analytic.

The term ‘symmetry’ is used by physicists in a way that goes beyond just the analytic symmetries. Crucially, there are bijections on the space of mathematical states that are afforded the epithet ‘symmetry’, even though they *are* taken to generate physical differences. An example is the group of Poincaré symmetries: the group of spatio-temporal displacements, rotations and boosts associated with special relativity. These symmetries certainly preserve *something*—in this case, the four-dimensional “distances” between spacetime events—but they must also be taken to do something *physical*. Otherwise, the existence of these symmetries could not be taken to be physically interesting, both in the sense of non-trivial and truly remarkable. And it surely *is* remarkable that particles with an average life-span (in their own reference frames) of microseconds manage to live long enough (according to *our* clocks) to reach detectors here on Earth: a fact explained by the Larmor time-dilation entailed by Lorentz symmetry and their near-light speeds.

We can accommodate these symmetries by relaxing our conditions on the analytic symmetries along two dimensions. First, we could try to limit the set of *quantities* we demand to be preserved. We still require the preserved quantities to be *physical* quantities (what would be the interest in preserving a non-physical quantity?), but we do not require the preservation of *all* of the physical quantities. Thus, unlike for the analytic symmetries, there will be some kinematically possible world in which some physical quantity changes its value under the symmetry. We set \mathcal{Q}' in equation (3) above to some proper subset of \mathcal{Q}_Φ (but not the empty set; otherwise we could not call the transformation a symmetry at all); which subset we choose will depend on the type of symmetries we are interested in. I propose to call these transformations *synthetic symmetries of the first kind*; call the set of these symmetries \mathcal{A}_1 (so ‘A’ does not stand for ‘analytic’!). Although I call these symmetries ‘synthetic’, their holding or not will, like the analytic symmetries, follow solely from the fact about which quantities we have decided to include in \mathcal{Q}_Φ —that is, solely in virtue of the representation relation that holds between the theory’s formalism and the physical properties and relations.

Alternatively, we could limit the *states* for which we demand preservation of values for the physical quantities. This is clearly equivalent to limiting the models in which the symmetries hold, to something short of all the kinematically possible worlds. In this case, equations (1)–(3) will fail. In particular, (3)’s failing makes it strictly incorrect to say that these transformations preserve \mathcal{Q}_Φ at all. Thus these transformations generate physical changes, just as the synthetic symmetries of the first kind do. However, we may still talk of a *restricted* preservation of *values* for \mathcal{Q}_Φ . I propose to call these *synthetic symmetries of the second kind*; call their set \mathcal{A}_2 . Unlike the analytic

¹⁶On the surprisingly thorough formal analogies between classical gauge systems and foreign-exchange markets, see Maldacena (2014).

symmetries and the synthetic symmetries of the first kind, we need to know *more* than just the theory’s interpretation to know whether a synthetic symmetry of the second kind holds. Specifically, we need enough contingent information to narrow down the kinematically possible worlds to those whose associated mathematical states are all ones for which the symmetry holds.

There are synthetic symmetries that are of the first, but not the second kind, since there are transformations which preserve the values of only a proper subset of \mathcal{Q}_Φ , but which do so in all kinematically possible worlds. I claim that Poincaré transformations in relativistic theories are of this kind, at least in most applications. For example, the mass and intrinsic spin of a relativistic quantum particle is preserved under Poincaré transformations, no matter what the state; but a Poincaré transformation may change a particle’s location, orientation or velocity.¹⁷

There are synthetic symmetries of the second, but not the first kind, since there are transformations which preserve the *values* of all physical quantities *in certain states*, but which nevertheless fail to preserve the physical quantities themselves. An example of such symmetries are provided by the permutations of bosons or fermions in many-particle quantum mechanics, *on an haecceitistic understanding* of the particles. In this theory, we restrict the space of states to either the totally symmetric (bosonic) or totally anti-symmetric (fermionic) states. On a haecceitistic interpretation, we have no grounds to demand that physical quantities be permutation-invariant. But no matter: *every* quantity is preserved under any permutation, since the bosonic or fermionic states themselves are permutation-invariant.

Finally, there are synthetic symmetries that are of both kinds at once (call their set \mathcal{A}_3), since there are transformations which preserve the values of only *some* of the physical quantities, in only *some* of the mathematical states.

The foregoing discussion of the three kinds of symmetry is summarised in Table 1 below (where k.p. stands for ‘kinematically possible’):

The synthetic symmetries (of either kind) do not form a group, since they do not include the identity transformation. However, the unions $\mathcal{A}_\Phi \cup \mathcal{A}_1$, $\mathcal{A}_\Phi \cup \mathcal{A}_2$ and $\mathcal{A}_\Phi \cup \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$ (corresponding in the table above to the first column, the first row, and all four quadrants, respectively) do all form groups, *so long as* we hold fixed (in the case of \mathcal{A}_1) the set of physical quantities and (in the case of \mathcal{A}_2) the set of mathematical states for which we demand the preservation of values. In fact, all three groups have \mathcal{A}_Φ as a normal subgroup.¹⁸ So we may form three factor groups. There are two good reasons for (contrary to what I have done) reserving the terms ‘synthetic of the first kind’, ‘synthetic of the second kind’ and ‘synthetic of both kinds’ for these factor groups: (1) they are groups; and (2) they do not involve “double-counting” the synthetic symmetries. The disadvantage is that the division of symmetries into analytic and synthetic would not be mutually exclusive, since the trivial symmetry, *id*, would count as both.

¹⁷Doesn’t it follow that I must be taking location, orientation and velocity as absolute quantities? No: the rest frame with respect to which these are defined may be set by another physical body—the fixed stars, for instance—which we have chosen not to explicitly represent in the formalism. Note that interpreting Poincaré transformations in this case as *analytic* symmetries would commit one to the absurd claim that the entire single-particle Hilbert space, which carries an irreducible representation of the Poincaré group, corresponds to just *one* physical state!

¹⁸A normal subgroup K of G is defined as follows: for all elements g in G , $gK = Kg$, or (equivalently) $gKg^{-1} = K$, where $gK := \{g \circ k \mid k \in K\}$, $Kg := \{k \circ g \mid k \in K\}$ and $gKg^{-1} := \{g \circ k \circ g^{-1} \mid k \in K\}$.

	Holds in all k.p. worlds <i>Necessary</i> $\mathcal{S}' = \mathcal{S}$	Fails in some k.p. world <i>Contingent</i> $\mathcal{S}' \subset \mathcal{S}$
Holds for all physical quantities <i>Indiscriminating</i> $\mathcal{Q}' = \mathcal{Q}_\Phi$	<i>Analytic</i> \mathcal{A}_Φ	<i>Synthetic 2</i> \mathcal{A}_2
Fails for some physical quantity <i>Discriminating</i> $\mathcal{Q}' \subset \mathcal{Q}_\Phi$	<i>Synthetic 1</i> \mathcal{A}_1	<i>Synthetic 3</i> \mathcal{A}_3

Table 1: Four kinds of symmetry

To expand on (2): since the analytic symmetries do not produce a physical difference, but may be formally non-trivial, there may be several synthetic symmetries which produce the *same* physical differences. For example, if a_1 is a non-trivial ($\neq id$) analytic symmetry, and a_2 is a synthetic symmetry, then $a_1^{-1} \circ a_2 \circ a_1$ will generate the same physical changes as a_2 . It therefore makes sense to represent the physical change in question not by a_2 but by the equivalence class $\{a^{-1} \circ a_2 \circ a \mid a \in \mathcal{A}_\Phi\}$. (These equivalence classes are precisely the elements of the factor groups just mentioned.) These factor groups remain non-trivial on the quotient space $\mathcal{S}/\mathcal{A}_\Phi$, further reflecting the fact that they are physically non-trivial.

To end this section, I will assimilate two types of physical symmetry, which are often discussed in the literature, into the framework just developed.¹⁹ Both apply to classical theories. The first type of symmetries are the *variational* symmetries, the group of transformations which preserve the Lagrangian function L . These will comprise both analytic symmetries and synthetic symmetries of the first kind. They cannot include synthetic symmetries of the second kind, since they are required to hold for all mathematical states, and thus all kinematically possible worlds. Whether they will comprise *all* of the analytic symmetries will depend on the details of the theory's interpretation, since L may or may not count as a physical quantity.²⁰ That they will comprise some synthetic symmetries is certain, since L will definitely not *exhaust* the physical quantities.

The second type of symmetries are the *dynamical* symmetries, which preserve the solutions of the dynamical equations. In my framework, this is tantamount to preserving the truth of T , the axioms of the theory. In systems subject to a variational treatment, this translates into preserving the fact that the Euler-Lagrange functions $E_\mu := \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^\mu} - \frac{\partial L}{\partial q^\mu}$ vanish. That is, the dynamical symmetries are those transformations which preserve the values of the quantities E_μ for those states in which $E_\mu = 0$. Clearly, therefore, the dynamical symmetries comprise *all three* kinds of symmetry. They include synthetic symmetries of the first and second kind, since their holding is restricted in terms of both quantities and states. The restriction of models to the dynamically

¹⁹Variational and dynamical symmetries are discussed in e.g. Wallace (2003).

²⁰Cf. footnote 7.

possible worlds is surely one of the most salient for synthetic symmetries of the second kind.

To sum up, a theory's symmetries may be categorised as analytic and synthetic: the analytic symmetries do not generate a physical difference, while the synthetic symmetries do. This difference is of the utmost importance, since we glean information from it about which elements of the theory's formalism are physically representative, and which are not. Therefore, identifying a theory's analytic symmetries is bound up with identifying its physical content.

3. A model of theory interpretation: the second triangle

3.1. *The need for interpretative constraints*

The interpretation of a physical theory involves the systematic pinning down of a representation relation. This relation holds between the mathematical elements immediately referred to by the theory's formalism—real-valued functions, Hermitian operators, manifold points, vector fields, and so on—and the putatively physical entities of which the theory is ultimately taken to be an account—as they might be: events, electrons, velocities, etc. There is plenty of room for disagreement here as to what 'the physical' comprises: it is not really a well-defined term, but rather a placeholder for whatever domain it is believed it is the business of physical theories generally to be about. Some countenance unobservable objects and properties (e.g. Armstrong (1978), Lewis (1983), Psillos (1995)); some are willing to go only as far as mind-independent "causal structure" (e.g. Ladyman 1998); others settle for just the appearances (e.g. van Fraassen (1980, 2008)).

Since at least the advent of the atomic theory and Mach's (1910/1992) philosophy of science, we are familiar with the predicament that this representation relation cannot be stipulated explicitly, at least for the unobservables. But let us first consider the prospects for explicit definition. Explicitly defined representation relations have an analogue in Davidsonian meaning-theories (e.g. Davidson 1973), in which an interpretation for an alien language is given as a finite set of principles that allows a recursive generation of biconditionals of the form ' S is true iff p ', where S is a sentence of the alien (object-) language, and p is a specification in the home (meta-) language of the truth-conditions of S . Similarly, in the case of physical theories, we might imagine interpretation to consist in something like the construction of a table with two columns. In the leftmost column we list certain mathematical items, and in the rightmost column we list their physical correlates: E represents energy, t represents time, and so on.

For the antecedently understood entities, the filling in of this table can mirror the Davidsonian approach. We do this by assuming that the theory is empirically adequate: i.e. we demand that the theory represent the physical world in such a way that it makes true claims about the observed phenomena. This requirement plays much the same role as Davidson's Principle of Charity. Here we may tentatively follow the experimentalist: for she knows how to use the theory to generate empirical predictions that she then uses to test the theory. (If we are interpreting a theory taken as empirically inadequate, we can still gain the needed information, since the experimentalist can tell us what results we would *expect* to find, according to the theory.)

We must remain tentative, however, since the experimentalist's identifications may rest on a lucky empirical regularity. The theory may contain covarying quantities, *all* of which are potential candidates for representing a particular physical quantity if any one of them is. If the experimentalist fixes on one of these candidates as the representative, we are permitted to remain agnostic

until further along the interpretative process, whether to follow the experimentalist or whether to opt instead for one of its covarying cousins. An example of this is provided by the Born rule in quantum mechanics. Experimentalists read probabilities for experimental outcomes straight off of the relative amplitudes, but a theoretician may insist that probabilities can only manifest as, for example, long-run relative frequencies, and so are represented elsewhere in the formalism. Nevertheless, it may be a result of the theory that, always or often, long-run relative frequencies covary with relative amplitudes. In such a case the experimentalist’s strategy would be vindicated in practice but not in principle.

We soon run into difficulties once we consider the unobservable entities that we take the theory to be about. Often a theory provides our *only* grasp on the physical entity being represented. (*Question*: What is spin? *Answer*: a quantum-mechanical degree of freedom whose state space supports finite-dimensional unitary representations of the group of rotations in three-dimensional Euclidean space.) This is certainly true of many unobservable physical entities, and it seems that theory goes at least some way in fixing the senses for terms which refer to observable entities too. (If we found that an object that ‘looked blue’, even in good conditions, did not reflect electromagnetic radiation in the appropriate wavelength interval, would we still describe it as blue? This gives us a reason in addition to the one just mentioned for being tentative in following the experimentalist’s operational interpretation.) So when it comes to filling in the rightmost column of the interpretation table, we find ourselves literally lost for words. And it would obviously be no advance to propose a new word. (If I ask you what **B** represents, and you tell me to call it ‘the magnetic field’, I still want to know what the magnetic field *is*.)

However, it may still be possible to define this part of the representation relation *implicitly*, given a substantial proviso. The proviso is that something like Lewis’s solution to Putnam’s paradox is correct, i.e. that there is a privileged collection of entities—objects, properties and relations—which are the preferred candidates for reference or representation.²¹ Once the candidates for representation are independently narrowed down, a unique maths-physics representation relation may be picked out by the meagre means available to us.

So we enter a second phase of the interpretative process. (A reminder: in the first phase we tentatively gave an explicit specification of the representation relation for the observable entities.) In this second phase we specify the *form* of the representation relation, e.g. whether it induces a one-one function from the mathematical states to possible physical states. (We have already seen, in Section 2, the link here with analytic symmetries.) And we may also specify “one end” of the relation’s extension where it links to the unobservable realm: namely, those elements of the mathematical formalism which straightforwardly represent the physical properties and relations. These elements are separated from the elements of the formalism which are mere descriptive artefacts, the elements of the theory that don’t represent anything. (The descriptive artefacts are the idle wheels of the theory, so far as representation of the physical goes; but they may be difficult or inconvenient to purge from the formalism.) In this way, the partial definition of the representation relation, combined with the claims of the theory (e.g. its equations) may suffice to select unique physical correlates for the representative mathematical elements.

²¹For a statement of Putnam’s paradox, see Putnam (1976, 1981) or Lewis (1984). For Lewis’s proposed solution, see Lewis (1984). For the inspiration for Lewis’s solution, see Carnap (1928/1967).

Without the Lewisian proviso, the second phase of interpretation is superfluous. For, no matter what formal constraints we place on the representation relation, without a sufficiently narrowed class of potential entities to represent we either cannot fail to find a representation relation that satisfies them, or cannot *but* fail. Failure is guaranteed if there is *nothing* at all to refer to, and success comes all too easy if there *are* things to refer to in sufficiently large numbers, but no objective divisions have been made between them. Guaranteed success and guaranteed failure are both disasters: neither could claim to be the pinning down of an interpretation.

Though substantial, the Lewisian proviso that there be objective unobservable structure ought to be granted by anyone who accepts both: (i) that they are engaging in an interpretative project that goes beyond the appearances; and (ii) that interpreting is a matter of fixing *representations*. Without (ii), it is hard to give a motivation to interpretation at all. Without (i)—i.e., without accepting that there is some objective unobservable structure to the world—there would be nothing for these interpretative projects to find out.

That is not quite right—we might instead adopt Lewis’s solution insincerely. Interpreting a theory does not mean treating it as true, it means finding out what the world would be like *if it were true*. The Lewisian machinery required for any theory’s being true can be taken just as hypothetically as the specific claims of any particular theory. If she is happy that the Lewisian proviso makes sense, the anti-realist may well be interested in literal interpretations of a physical theory, just as a realist may be interested in interpreting a theory they know to be false.

The puzzle remains how to decide which constraints to lay down to implicitly define the representation relation. That is the puzzle I wish here to address. The obvious, quasi-Davidsonian, recommendation—to maximise the true claims of the theory—is no help once we have entered the second phase of interpretation, since: (i) we may not be trying to interpret a theory which is taken to be true (we may be seeking interpretative clarification for a long-since-falsified theory); and (ii) if the theory is taken to be true, then our interpreting it—i.e. our laying down a representation relation for it—is the very project of finding out the true claims that are to be maximised. It is no advance to know that we are trying to maximise the theory’s true claims until we know what the true claims are, i.e. until we have given a theory an interpretation.

Thus we have a second triangle, the vertices of which mutually determine one another: the theory’s formalism, the physical world, and the representation relation that is taken to hold between them.

3.2. A proposal

The natural weakening of the quasi-Davidsonian recommendation—to maximise empirical adequacy—is no help in the second phase either, since maximising empirical adequacy only gets us as far as the first phase. We are in need of more constraints. But the triangle of theory, world and interpretation is connected to the triangle of states, quantities and symmetries. My proposal is to use this connection to provide those constraints.

Recall (from equation (4), above) that which symmetries are analytic is determined by which quantities are physical. Important for us now is that the reverse is also true: the analytic symmetries determine the physical quantities. That is, the analytic symmetries determine which mathematical functions in the theory’s formalism are the representatives of the physical properties and relations. We have also seen that the analytic symmetries determine the form of the representation relation

between mathematical states and physical states: i.e. they determine a partition of the space \mathcal{S} of mathematical states into equivalence classes, each one of which is the representative for a single physical state. Therefore the analytic symmetries \mathcal{A}_Φ determine the “mathematical end” of the representation relation, as required for its implicit definition. The natural divisions in the physical world then do the rest (we hope). We can then glean an understanding of what the theory says the world is like by studying the quotient space $\mathcal{S}/\mathcal{A}_\Phi$ and its features, safe in the knowledge that the redundant structure has been purged from the formalism.

In this way, analytic symmetries serve the role of Reichenbach’s (1958, 1965) constitutive *a priori* principles: as constraints on our theory’s representation relation, they are a precondition on the theory’s formalism having empirical significance. As such, they cannot themselves be taken as empirical claims; rather they are the mediators for the encounter between the mathematical formalism and the physical world.

But surely (you may be thinking) we have made no advance. For how are we supposed to know, before we have laid down a representation relation, *which* symmetries are the analytic ones? Surely we first need to know which quantities are the physical quantities, and that requires knowing the representation relation we are trying to pin down.

That is not quite right. For there are ways to make sensible decisions about which symmetries are analytic purely on the basis of formal considerations (i.e. divorced from what represents what) and on the results of the first phase of interpretation, in which representatives for the observable entities were tentatively identified. In short, my proposal is to *maximise the analytic symmetries, subject to empirical adequacy*.

The proposal runs as follows. The first phase of interpretation lays down a lower limit for what changes count as physical: if a transformation effects an empirically observable change, then it must be physical. Failure to respect this lower limit will lead to the theory being empirically inadequate. That is why, for example, it would be a disaster to treat the dynamical symmetries (mentioned in Section 2, above) as exclusively analytic, for in doing so we would be taking *all* of the solutions to the theory’s equations as representing the *same* physical world: namely, the actual world. Yet our experimental procedures make conceptual room for finding the world to be other than it is, but for it still to be a solution to the theory’s equations. That is, we can say in purely *observational* terms what would have been different in these unactualised, dynamically possible worlds. For example, we can say what the world would have looked like if it had had *different* boundary conditions but the *same* physical laws. (This goes to show that empirical adequacy is a more complicated matter than simply representing the actual phenomena.)

However, keeping fixed that our theory is empirically adequate, we may still find symmetries of the theory which generate no observable difference. In this case, there would be no argument on the basis of preserving empirical adequacy for not taking it to be an analytic symmetry. And, lacking a full interpretation for the theory, we have no other arguments to go on.²²

The assimilation of symmetries previously thought to be synthetic to the analytic symmetries can go one of two ways, corresponding to the two kinds of synthetic symmetry. Each has its own

²²For special cases of support for the view that the maximisation of analytic symmetries subject to empirical adequacy is always an interpretatively *possible* strategy, see Hacking (1977) and Belot (2001).

implications for the physical properties and relations. In assimilating a symmetry formerly thought to be synthetic of the first kind to the analytic symmetries, we thereby reduce the set of quantities we take to be physical, in virtue of equation (4), above. This induces a corresponding reduction in the natural properties and relations represented by the theory to compose the physical world.

For example, consider permutation symmetry in a classical mechanical theory describing point masses moving under mutual forces. Any transformation which permutes equally massive bodies will either be an analytic symmetry or a synthetic symmetry of the first kind. Suppose that, according to the theory's laws, observable phenomena will not distinguish between mathematical states in which point masses possessing the same mass are transposed. Therefore, our recommendation is to treat the permutations as analytic symmetries. The effect is to treat each particle's identity, divorced from its physical properties like mass, as redundant formal structure. Thus our interpretative methodology recommends anti-haecceitism, in Lewis's (1986) sense.

In assimilating a symmetry formerly thought to be synthetic of the second kind to the analytic symmetries, we thereby reduce the space of mathematical states (and the space of kinematically possible worlds) we take to represent any physical state at all, in virtue of equation (2), above. This induces a reduction in the natural properties and relations, since the amount of worlds is decreased by which we may distinguish them.

For example, consider the Galilean symmetries, again for a system of classical point masses. This is the group of translations, rotations, and non-relativistic boosts acting on the centre of mass of the system which do not affect the relative distances and velocities between the masses. In this system, physical changes are only observed for the relative quantities, so we are led to treat the Galilean symmetries as analytic. Now the Galilean symmetries were originally synthetic symmetries of the first kind, but there is a set of related transformations which are synthetic of both kinds. These are the angular boosts: they change the angular momentum of the system. However, angular momentum is preserved under dynamical evolution, so according to the laws of the theory we would never *observe* an angular boost. We might therefore try to strike the system's angular momentum from the register of physical quantities.

This is a riskier strategy than for synthetic symmetries of the first kind, since we can already make observational sense of this move being the wrong one: i.e. we can say in observational terms what the world would have been like had its angular momentum been different.²³ However, once fixed the angular momentum of the system is fixed for all time, so there is no danger of us being wrong in the future, so long as we set it initially to the correct value. If it were dropped from the physical quantities, there would be the remaining mystery of the presence of a seemingly kinematical (i.e. invariant) parameter, in the units of angular momentum, in the laws of the theory (cf. Belot 2003). The continuing presence of this parameter makes a sham out of the attempt to repudiate universal rotation. But still we might be lucky: if its value is zero, then it will *not* appear in the laws, once they are written in terms of the purely relational quantities. Thus universal rotation can be satisfactorily eliminated so long as we are "lucky enough" to live in a universe that does not rotate. (Of course, once eliminated, it will not appear as a matter of *luck*, since non-rotating universes will exhaust the kinematically possible worlds.)

²³For a compelling description of this situation, see Poincaré (1952).

To continue the parallels with Reichenbach mentioned above, I note that the maximisation of the analytic symmetries (subject to empirical adequacy) is something of a “meta-conventional principle”: a conventional principle that applies to the setting down of conventional principles (namely, the analytic symmetries themselves). It plays the role of Reichenbach’s (1958, 1965) regulative *a priori* principles, such as the elimination of universal forces, or minimisation of cosmic coincidences.

It is worth pointing out that the proposal outlined above for eliminating redundant representational structure may seem to pit the philosopher against the physicist. For redundant structure often provides convenience for calculation—it may even be essential for the tractability of some problems in cases where no redundancy-free techniques are known.²⁴ An important class of examples are provided by gauge theories. For example, the electromagnetic four-vector potential features in the standard Lagrangian density for a charged field, even though gauge transformations, which are quasi-symmetries of the Lagrangian density, are standardly interpreted as analytic symmetries in the sense of this paper.²⁵ However, there is no need to see the philosopher and physicist as opponents here, since their different sensibilities are motivated by different aims. It would certainly be improper for the philosopher to hamstring physics practice by some ascetic demand to use only as much mathematical structure as could be accounted for in the corresponding physical ontology. But conversely, the physicist need not be offended by the philosopher’s insensitivity to calculational techniques in the pursuit of the best interpretation of a physical theory.

4. Conclusion

I have argued that a theory may be given a unique literal interpretation, even for unobservable entities, given the imposition of the right constraints: namely, the elimination of distinctions without a difference. We have seen that this is because the representation relation linking the theory to the physical world is tied to the group of symmetries that do not generate a physical difference. We proceed in two phases. During the first phase we set up representational links between the theory and the observable portion of the physical world, under the assumption that the theory is empirically adequate (or similar). In the second phase, we maximise the theory’s analytic symmetries, taking advantage of the representational links forged in the first phase so as not to compromise empirical adequacy. The result is an interpretation for the theory that prompts appropriate reform towards a *new* formalism, in which the physical properties and relations—including the unobservable ones—are transparently represented without redundancy.

Finally, it might be objected that my proposal amounts to nothing more than a form of verificationism, the doctrine that only observable differences make for meaningful differences. But that would not be an accurate characterisation of my proposal. Like Leibniz, I do not claim that non-observable physical difference do not make *sense*; my claim is that the maximisation of analytic symmetries serves as a methodological principle, or a “super-empirical virtue”—a means to overcome theoretical under-determination, to get at an account of what the world is like.

²⁴I am grateful to an anonymous referee for making this observation.

²⁵A *quasi-symmetry* is a transformation which preserves the Lagrangian density up to a divergence term; see Brading & Brown (2003, pp. 93-95).

As with any methodological principle, we cannot say *why* we should expect the world to be so kind as to oblige us in vindicating it. For all we know, the world is awash with empirically inert structure. We can only say that it provides us with a means to go on, and hope for the best. (The observable/unobservable distinction is continually shifting, so we might even hope to *find out* that we were wrong.) All that is true unless we identify the two—i.e. unless we take it as *constitutive* of truth that these human-made constraints be satisfied, at least in the long run. I suppose that would constitute a form of pragmatism. A proper treatment of its cogency will have to be saved for another time.

Acknowledgements

Many thanks to Jeremy Butterfield for many conversations; also to Neil Dewar, Thomas Moller-Nielsen and audiences in Bristol. I am also grateful for generous support from the British Academy.

References

- Alexander, H. G. (ed.) (1998), *The Leibniz-Clarke Correspondence*, Manchester: Manchester University Press.
- Armstrong, D. M. (1978), *Universals and Scientific Realism, Volume 2*. Cambridge: Cambridge University Press.
- Belot, G. (2001), ‘The Principle of Sufficient Reason’, *The Journal of Philosophy*, **98** (2), pp. 55-74.
- Belot, G. (2003), ‘Notes on symmetries’, in Brading and Castellani (2003), pp. 393-412.
- Blackmore, J. (ed.) (1992), *Ernst Mach: A Deeper Look*. Boston Studies in the Philosophy of Science, vol. 143. Dordrecht: Kluwer Academic Publishers.
- Brading, K. and Brown, H. R. (2003), ‘Symmetries and Noether’s theorems’, in Brading & Castellani (2003), pp. 89-109.
- Brading, K. and Castellani, E. (eds.) (2003), *Symmetries in Physics: Philosophical Reflections*, Cambridge: Cambridge University Press.
- Broyles, A. A. (1970), ‘Space-time position operators’, *Physical Review D* **1**, pp. 979-988.
- Butterfield, J. N. (1989): ‘The Hole Truth’, *British Journal for the Philosophy of Science* **40**, pp. 1-28.
- Butterfield, J. N. (2006), ‘On symplectic reduction in classical mechanics’, in J. Butterfield and J. Earman (eds.), *The Handbook of Philosophy of Physics*, North Holland, pp. 1-131.
- Carnap, R. (1928/1967), *Der Logische Aufbau der Welt*, translated by R. George (1967): *The Logical Structure of the World*. Berkeley: University of California Press.
- Carnap, R. (1956), *Meaning and Necessity*. Chicago: University of Chicago Press.
- Carnap, R. (1966), *The Philosophical Foundations of Physics*. New York: Basic Books.

- Caulton & Butterfield (2012), ‘Symmetries and Paraparticles as a Motivation for Structuralism’, *British Journal for the Philosophy of Science* **63**, pp. 233-85.
- Davidson, D. (1973), ‘Radical Interpretation’, *Dialectica* **27**, pp. 314-328.
- Earman, J. and Norton, J. D. (1987), ‘What Price Spacetime Substantivalism’, *British Journal for the Philosophy of Science*, **38**, pp. 515-25.
- French, S. and Ladyman, J. (1999), ‘Reinflating the semantic approach’, *International Studies in the Philosophy of Science*, **13**(2), pp. 103-121.
- French, S. and Redhead, M. (1988), ‘Quantum physics and the identity of indiscernibles’, *British Journal for the Philosophy of Science* **39**, pp. 233-46.
- Frigg, R. and Hartmann, S. (2008), ‘Models in science’, in Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Stanford University, Stanford.
- Hacking, I. (1977), ‘The Identity of Indiscernibles’, *Journal of Philosophy* **72**, pp. 249-56.
- Henneaux, M. & Teitelboim, C. (1992), *Quantization of Gauge Systems*. Princeton, NJ: Princeton University Press.
- Isham, C. J. (1999), *Modern Differential Geometry for Physicists*, Second Edition. Singapore: World Scientific.
- Johnson, J. E. (1969), ‘Position operators and proper time in relativistic quantum mechanics’, *Physical Review* **181**, pp. 1755-1764.
- Ladyman (1998), ‘Structural realism: epistemology or metaphysics?’, *Studies in the History and Philosophy of Science*, **29A**, pp. 409-24.
- Lewis, D. (1970a), ‘How to Define Theoretical Terms’, *Journal of Philosophy* **67**, pp. 427-46.
- Lewis, D. (1970b), ‘General Semantics’, *Synthese* **22**, pp. 18-67.
- Lewis, D. (1980), ‘Index, Context, and Content’, in S. Kanger & S. Öhman (eds.), *Philosophy and Grammar* (Dordrecht: Reidel).
- Lewis, D. (1983), ‘New Work for a Theory of Universals’, *The Australasian Journal of Philosophy* **61**, pp. 343-77.
- Lewis, D. (1984), ‘Putnam’s Paradox’, *Australasian Journal of Philosophy* **62**, pp. 221-36.
- Lewis, D. (1986), *On the Plurality of Worlds*. Oxford: Blackwell.
- Lutz, S. (2010), ‘On a Straw Man in the Philosophy of Science: A Defense of the Received View’. Draft preprint.
- Mach, E. (1910/1992), ‘Sensory elements and scientific concepts’, translation by J. Blackmore of ‘Sinnliche Elemente und naturwissenschaftliche Begriffe’, *Archiv für Physiologie*, **136** (1910), pp. 263-74; in Blackmore (1992).
- Maldacena, J. (2014), ‘The symmetry and simplicity of the laws of physics and the Higgs boson’. ArXiv:1410.6753.
- Messiah, A. and Greenberg, O. W. (1964), ‘The symmetrization postulate and its experimental foundation’, *Physical Review* **136** (1B), pp. 248-67.

- Montague, R. (1970), 'Universal grammar', *Theoria* **36**, pp. 373-398.
- Poincaré, H. (1952), *Science and Hypothesis*. Translated by W. J. Greenstreet. New York: Dover Publications.
- Psillos, S. (1995), 'Is structural realism the best of both worlds?', *Dialectica* **49**, pp. 15-46.
- Putnam, H. (1976), 'Realism and reason', Presidential Address to the Eastern Division of the American Philosophical Association, December 1976. *Proceedings and Addresses of the American Philosophical Association* **50**, pp. 483-498, reprinted in Putnam (1978), pp. 123-40.
- Putnam, H. (1978), *Meaning and the Moral Sciences*. New York: Routledge.
- Putnam, H. (1981), *Reason, Truth and History*. Cambridge, Mass.: Cambridge University Press.
- Quine, W. V. O. (1948), 'On What There Is', *The Review of Metaphysics*, Vol. 2, No. 5, pp. 21-38.
- Quine, W. V. O. (1960), *Word and Object*. Cambridge: Harvard University Press.
- Reichenbach, H. (1958), *The Philosophy of Space and Time*. Translation, by Maria Reichenbach and John Freund, of H. Reichenbach, *Philosophie der Raum-Zeit-Lehre*. Berlin/Leipzig: Walter de Gruyter 1928. New York: Dover Publications.
- Reichenbach, H. (1965), *The Theory of Relativity and A Priori Knowledge*. Translated by Maria Reichenbach. Berkeley: University of California Press.
- Rynasiewicz, R. (1994), 'The lessons of the hole argument', *The British Journal for the Philosophy of Science*, **45**, pp. 407-36.
- Stalnaker, R. (1976), 'Possible Worlds', *Nous*, **10**, pp. 65-75.
- Stalnaker, R. (1979), 'Anti-essentialism', *Midwest Studies in Philosophy* **4**, pp. 343-55.
- Suppe, F. (ed.) (1974), *The Structure of Scientific Theories*. Urbana, IL: University of Illinois Press.
- Tung, W. (1985), *Group Theory in Physics*. Singapore: World Scientific Publishing.
- van Fraassen, B. C. (1967), 'Meaning relations among predicates', *Nous* **1**, pp. 161-79.
- van Fraassen, B. C. (1980), *The Scientific Image*. Oxford: Oxford University Press.
- van Fraassen, B. C. (2000), 'The semantic approach to scientific theories', in Sklar, L. (ed.), *Philosophy of Science volume 2: The Nature of Scientific Theory*, pp. 175-194. Garland Publishing, Inc.: New York.
- van Fraassen, B. C. (2008), *Scientific Representation: Paradoxes of Perspective*. Oxford: Oxford University Press.
- Wallace, D. (2003), 'Time-dependent symmetries: the link between gauge symmetries and indeterminism', in Brading & Castellani (eds.) (2003), pp. 163-173.
- Weatherall, J. O. (2015), 'Regarding the hole argument', forthcoming in *the British Journal for the Philosophy of Science*.