# 4

# Philosophy of Symmetry

***Précis.*** *On the Representation View, time reversal has direct empirical significance, breaks the symmetry-to-reality inference, and can be violated in a time-symmetric spacetime.*

Modern physics is firmly rooted in symmetry. Some twentieth-century physicists came reluctantly to appreciate this, with Wigner (1981) recalling how Schrödinger wanted to see the "Gruppenpest" abolished. However, this was largely due to the group theory formalism being viewed at the time as arcane.[1] In more general terms, symmetry has been a cornerstone of physics at least since the eighteenth-century correspondence of Leibniz and Clarke (2000).

Nevertheless, interpreting symmetry is a subtle business. A symmetry is in one sense a relation between two descriptions and in another sense a statement that they are one and the same. Hans Halvorson[2] illustrated the tension using this charming example:

From the perspective of group theory, suppose I ask, "How many groups of order two are there?" You should say: one. But now suppose I ask, "How many groups of order two are there among the subgroups of the Klein four-group?" And you should of course say: three!

[1] MIT physicist John Slater seems to have been particularly scandalised: "The authors of the 'Gruppenpest' wrote papers which were incomprehensible to those like me who had not studied group theory. . . . The practical consequences appeared to be negligible, but everyone felt that to be in the mainstream one had to learn about it. Yet there were no good texts from which one could learn group theory. It was a frustrating experience, worthy of the name of a pest. I had what I can only describe as a feeling of outrage at the turn which the subject had taken" (Slater 1975, pp.60–2).

[2] Halvorson (2019, p.259). Halvorson attributes the example to John Burgess, though in a different context.

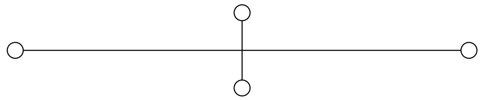

Figure 4.1 The Klein four-group $\mathbb{V}$ describes the symmetries of the figure above: the identity, a flip about each axis, and a flip about both.

This example hearkens back to an old (Fregean) debate.[3] One way to clarify what's going on is to observe that although there is only one group of order two, there are three representations of it on the Klein four-group (Figure 4.1), in the sense of three distinct isomorphisms[4] $\varphi_p, \varphi_t, \varphi_{pt}$ from $\mathbb{Z}_2$ into $\mathbb{V}$. In short, we can dissolve the puzzle by cleanly separating two concepts: the two-element group and its representations in another structure, the Klein four-group.

The Representation View of symmetries developed in Chapter 2 follows a similar strategy. The proposal is to cleanly separate the concept of a symmetry into two parts: a symmetry of spacetime and its representation on a state space. The thesis of this chapter is that a number of issues in the philosophy of symmetry can be clarified and improved by adopting this perspective.

I begin in Section 4.1 by proposing an interpretation of what it means to be a 'dynamical' symmetry, including time reversal symmetry, which recovers its standard definition in the literature as a transformation that 'preserves' the solution space of a dynamical equation. I will then argue that the Representation View has implications for three related philosophical discussions and show how time reversal appears in each.

The first discussion is about epistemology (Section 4.3): if two descriptions related by a dynamical symmetry are empirically indistinguishable, then how can we come to know about that symmetry? Following Kosso (2000) and Brading and Brown (2004), the literature about this question has explored how to answer this question by separating a subsystem with a symmetry from its environment, which does not generally have that symmetry: they call this the 'direct' empirical significance of a symmetry. I will introduce this idea in Section 4.3 and argue that both time translations and time reversal symmetry have direct empirical significance in this sense.

[3] Frege (1892) discusses a similar issue in his famous analysis of the meaning of '='. Another influential discussion of mathematical equivalence was given by the Bourbaki group; see Burgess (2015, Chapter 3) for a review.

[4] Writing $\mathbb{Z}_2 = \{0, 1\}$ (with identity 0) and the Klein four-group as $\mathbb{V} = \{1, p, t, pt\}$ (with identity 1), they are defined by $\varphi_p(1) = p$, $\varphi_t(1) = t$, $\varphi_{pt}(1) = pt$, and $\varphi_n(0) = 1$ for each $n = p, t, pt$.

https://doi.org/10.1017/9781009122139.005 Published online by Cambridge University Press

The second discussion is about reference (Section 4.4): if two mathematical models are related by a dynamical symmetry, does it follow that they must refer to the same physical situation? In the debate between Leibniz and Clarke (2000), Leibniz seemed to think so, and this question has recently been the subject of much debate. I will refer to this as the 'symmetry-to-reality inference', following Dasgupta (2016), and argue in Section 4.4 that there are a number of ways in which it can fail.

The final discussion is about whether a dynamical symmetry is necessarily a spacetime symmetry, and vice versa (Section 4.5), as set out in a well-known pair of symmetry principles (SP1) and (SP2) by Earman (1989, §3.4). In Section 4.5, I will argue that on the Representation View, there are senses in which both can fail.

## 4.1 Representing Dynamical Symmetries

### *4.1.1 What Is a Dynamical Symmetry?*

Standard practice has it that a dynamical symmetry is a state space automorphism that preserves the solution space of a dynamical equation.[5] Let me begin by heading off some confusion with a few words about what a dynamical symmetry is not.

First, some discussions define dynamical symmetry as preserving the 'form' of a dynamical equation, which Belot (2003) calls a *nomenclature* symmetry and others call *passive* in analogy to passive and active spacetime symmetries.[6] This deserves the warning rumoured to have been issued by mathematician Graeme Segal: "I defy anybody not to get confused about the difference between passive and active".[7] For example, one must be careful that such symmetries are not viewed as transforming $F = ma$ in such a way that acceleration goes to velocity. To avoid this, the nomenclature approach requires careful analysis of the syntactic 'form' and logical structures of the sentences of a theory.[8] As Belot points out, this often amounts to the same thing in practice. However, these technicalities take one too far afield, and so I will avoid the passive-nomenclature approach.

5 Cf. Belot (2003, 2007, 2013), Butterfield (2006c, 2021), Caulton (2015), De Haro and Butterfield (2019), Earman (1989, §3.4), Møller-Nielsen (2017), Read and Møller-Nielsen (2020), Uffink (2001, p.314), and Wallace (2022), among many others.

6 The 'passive' usage can be found, for example, in Brown and Holland (1999), Redhead (2003, §3), and Brading and Castellani (2007, §4.1).

7 I learned this from Jeremy Butterfield (personal correspondence).

8 The efforts of Barrett and Halvorson (2016a,b), Dewar (2022), Halvorson (2012, 2013, 2019) and collaborators, inspired by Glymour (1970, 1980), provide an enlightened modern expression of this approach, where it is usually referred to as 'translation between theories'.

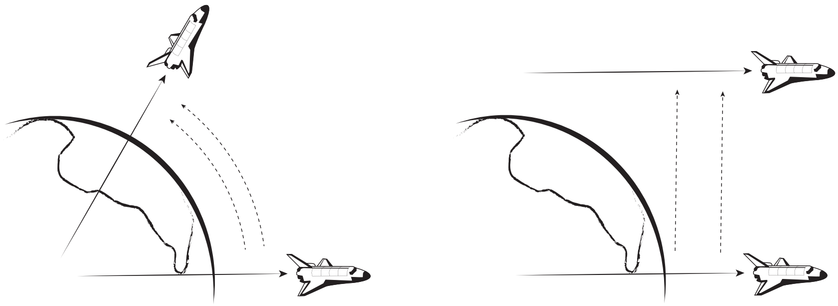

Figure 4.2 In a dynamical equation with a spherically symmetric potential, rotation (left) is a dynamical symmetry while translation (right) is not.

Second, dynamical symmetries are usually more restrictive than 'kinematical' symmetries, by which I mean the automorphisms of a state space. For example, in Hamiltonian mechanics, the kinematical symmetries are the symplectic transformations (see Section 3.3.1). But, when the state space is equipped with a dynamics, such as a Hamiltonian describing a spherically symmetric potential, then only some of the kinematical symmetries preserve the dynamics, such as the rotations about the centre of the potential (see Figure 4.2). Nor do dynamical symmetries necessarily leave an individual solution invariant, as Belot (2003, §4.2) is careful to note: although dynamical symmetries leave the space of all solutions invariant, they often transform each individual solution to a different one, as when rotation transforms a body moving east to a body moving north.

There is a puzzle about the meaning of dynamical symmetries that is not widely appreciated, which is the question of what makes an equation 'dynamical'. One might be tempted to answer that it is the usage of the term that does so: if an equation refers to change over time, then the equation is dynamical. But, what aspect of the formalism distinguishes change with respect to time from change with respect to anything else, like space?

For example, consider a bead on a string in Hamiltonian mechanics, with position and momentum represented by the coordinates $(q, p) \in \mathbb{R} \times \mathbb{R}$. One might wish to declare that motion is described by the solutions to Hamilton's equation, for some smooth function $h(q, p)$:

$$\frac{d}{dt}q(t) = \frac{\partial h}{\partial p}, \qquad \frac{d}{dt}p(t) = -\frac{\partial h}{\partial q}. \tag{4.1}$$

But, by choosing the function $h = p$, we find the solutions describe exactly the spatial translations, $q(t) = q_0 + t$ and $p(t) = p_0$, for each initial state $(q_0, p_0)$. In other words, the equation that was declared to be dynamical has

https://doi.org/10.1017/9781009122139.005 Published online by Cambridge University Press

the very same structure as one that can be used to refer to non-dynamical translations in space. What distinguishes these two situations?

The problem reappears in field theory as well. In his discussion of classical dynamical symmetry, Earman (1989, pp.45–6) considers any theory with a set of models each of the form $(M, A_1, A_2, \ldots, P_1, P_2, \ldots)$, where $M$ is a smooth manifold, each $A_i$ is an 'absolute object' field on $M$ characterising spacetime structure, and each $P_i$ is a 'dynamic object' field on $M$ characterising physical matter-fields. A dynamical symmetry is then defined to be a transformation of the dynamic objects $P_i$ that preserves the theory's set of models. Here the question becomes the following: What justifies calling a given field 'dynamic' as opposed to 'absolute'?

My answer is to propose the Representation View: a symmetry of a state space can be interpreted as a 'spacetime symmetry' only if it is an element of a *representation* of a spacetime symmetry structure, as argued in Section 2.3. So, in particular, a symmetry can only be viewed as dynamical in the sense of describing 'time translation' if it is a representation of time translations. This is what it means when one says that a curve through state space represents 'time evolution' or that a dynamical field on spacetime is 'dynamical'.[9] Let me now make the consequences of this more precise.

### *4.1.2 Temporal and Dynamical Symmetries*

Suppose we have on the one hand a group of symmetries $G$, which could include spacetime or even gauge symmetries, with a preferred subgroup $\mathbb{T} \subseteq G$ interpreted as 'time translations'. In earlier discussions we often focused on $\mathbb{T} = (\mathbb{R}, +)$ to represent time translations; now, I would like to proceed a little more generally and allow $\mathbb{T}$ to be whatever choice of time translations we happen to find appropriate. We will see two alternatives below.

On the other hand, suppose we have a structure $M$ interpreted as 'state space' and which has a group of automorphisms $\mathrm{Aut}(M)$. In earlier discussions, we took $M$ to be a Hilbert space for quantum mechanics, or a symplectic manifold for classical mechanics, but we will leave this open for now. The central postulate of the Representation View is that if the group of symmetries $G$ has meaning in state space, it is through the existence of a representation $\varphi : G \to \mathrm{Aut}(M)$, which is a 'homomorphic copy' of $G$ amongst the symmetries of $M$. This explains why a given state space automorphism $S = \varphi(g)$

[9] A similar view is advocated by Belot (2007, p.171): "Time, in one of its facets, is represented in this scheme by ... $\mathbb{R}$-actions", where Belot uses '$\mathbb{R}$-action' to refer to what I call a state space representation of the time translation group $(\mathbb{R}, +)$.

is interpreted in the same way as a group element $g$: time translation, rotation, gauge, or any of the like. It is because this is exactly what it represents.

This cleanly separates the concept of a symmetry into two parts, related by a representation: a spacetime part and a state space part. It also allows the concept of a 'dynamical symmetry' to be separated into two parts, corresponding to the two 'sides' of the representation relation. To distinguish them, I will refer to the spacetime version as 'temporal symmetries' and the state space version as 'dynamical symmetries'. The former can be defined in short as transformations that preserve time translations:

> **Definition 4.1** (temporal symmetry) Given a group $G$ with a subgroup $\mathbb{T} \subseteq G$ interpreted as 'time translations', a *temporal symmetry* $g \in G$ is an element that acts invariantly on $\mathbb{T}$ by the automorphism $\alpha_g(\cdot) = g(\cdot)g^{-1}$, in that $\alpha_g(t) \in \mathbb{T}$ for all $t \in \mathbb{T}$. If every element of $G$ is a temporal symmetry, then $G$ is called a *temporal symmetry group*.

The idea is, in short, that a temporal symmetry is a symmetry that preserves time translations, and we refer to a group of these symmetries as a temporal symmetry group. We have seen one example of this already in Section 2.6: let the group of time translations be $\mathbb{T} = (\mathbb{R}, +)$, and let $G := (\mathbb{R}, +) \rtimes \{\iota, \tau\}$ be its extension to include time reversal. Then, since every element $g \in G$ maps each time translation to another time translation, $gtg^{-1} \in (\mathbb{R}, +)$ for all time translations $t$, we say that $G$ is a temporal symmetry group for $(\mathbb{R}, +)$.

Here is a more interesting example. Let $G = \mathcal{P}$ be the Poincaré group. Then, by definition, each spatial translation $s \in \mathcal{P}$ maps each time translation $t$ to itself via the automorphism $\alpha_s(t) := sts^{-1} = t$. So, the spatial translations preserve $\mathbb{T}$, which means they are temporal symmetries. The Lorentz boosts are a little more subtle: we need to distinguish two cases, illustrated in Figure 4.3.

1. In some applications, we may choose a fixed inertial reference frame, in which time translations are given by a one-parameter group isomorphic to $\mathbb{T} = (\mathbb{R}, +)$. The Lorentz boosts are not temporal symmetries given this definition of time translations, since they result in a different reference frame and therefore a different flow of time translations.
2. In other applications, we might define time translations to be the larger group $\mathbb{T}$ of all the translations along any timelike line (a four-parameter Lie group). Then Lorentz boosts are indeed temporal symmetries, since boosts preserve the set of all timelike vector fields.[10]

[10] More formally, in the first case we define the group of time translations $\mathbb{T}$ as the flow along a fixed timelike vector field $\xi^a$ that satisfies the geodesic equation ($\nabla^b \xi_b \xi^a = \mathbf{0}$) in Minkowski spacetime.

https://doi.org/10.1017/9781009122139.005 Published online by Cambridge University Press

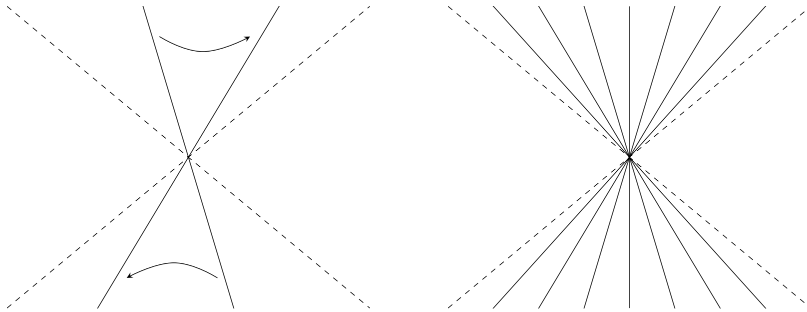

Figure 4.3 Boosts are not a temporal symmetry of translations along a single timelike line (left) but are of the group of all timelike translations (right).

Thus, whether a transformation is a temporal symmetry depends on how we define time translations, as one might expect.

Careful attention to this aspect of the Representation View can help to avoid confusion: for example, as Belot (2013, pp.327,331) has emphasised, boosts do not generally preserve the Euler–Lagrange equations or Hamilton's equations. This led Baker[11] to suggest that Lagrangian and Hamiltonian mechanics leave some puzzlement as to the origin of absolute velocity. But, from the present perspective, Belot's observation is just an instance of the first case above: a preferred frame of reference is needed whenever we choose an Euler–Lagrange or Hamiltonian representation of the time translation group $(\mathbb{R}, +)$. The second case can be recovered by instead representing time translations by a collection of Hamiltonian (or Lagrangian) descriptions related by boosts. Whether boosts are a temporal symmetry in either theory just depends on our definition of time translations.

Temporal symmetries can thus be understood entirely at the level of group transformations. To define dynamical symmetries, we must now consider the other side of the Representation View: a representation on state space. The definition of a dynamical symmetry is then natural, in that it is simply a 'representation' of a temporal symmetry:

This vector field (and hence the group of time translations) is transformed non-trivially by a Lorentz boost $\xi^a \mapsto \Lambda^b_a \xi^a \neq \xi^b$, and so it is not a temporal symmetry. The second case defines time translations to be the group $\mathbb{T}$ of flows along *any* of the set of timelike geodesic vector fields $\{\xi^a \mid \xi^a \xi_a > 0, \nabla^b \xi_b \xi^a = \mathbf{0}\}$. This set is preserved by Lorentz boosts, and so boosts are a temporal symmetry of these time translations.

[11] Baker (2020). "What are symmetries?" In: Unpublished Manuscript, http://philsci-archive.pitt.edu/18524/

https://doi.org/10.1017/9781009122139.005 Published online by Cambridge University Press

**Definition 4.2** (dynamical symmetry) Let $\mathbb{T} \subseteq G$ be a group of time translations. Given a representation of time translations $\varphi : \mathbb{T} \to \mathrm{Aut}(M)$, we say that $G$ is a group of *dynamical symmetries* if and only if (1) $G$ is a temporal symmetry group; and (2) there is a non-trivial extension of $\varphi$ to all of $G$, i.e. a representation $\tilde{\varphi} : G \to \mathrm{Aut}(M)$ such that $\tilde{\varphi}(t) = \varphi(t)$ for all $t \in \mathbb{T}$.

The first part of this definition just says that dynamical symmetries are temporal symmetries: they preserve the structure of time translations. The second part expresses a consequence of the Representation View: a dynamical symmetry on state space is just the representative of a spacetime symmetry $g \in G$ in some representation. When a group element $s$ can be represented as a dynamical symmetry in this way, it is common in physics to call this *g-invariance*, using phrases like rotational invariance, translation invariance, time reversal invariance, and so on.

To take an example from quantum theory, let $U : t \mapsto U_t$ be a representation of the time translations $(\mathbb{R}, +)$ amongst the unitary (and antiunitary) operators on a Hilbert space.[12] Let $G$ be the extension of this group to include time reversal $\tau t \tau^{-1} = -t$. As we have seen above, this group $G$ is a temporal symmetry group. Suppose now that we have an extension of $U$ to all of $G$, and let $T := U_\tau$ be the representative of time reversal. Then, $G$ is a dynamical symmetry. In particular, the fact that a representation is a homomorphism implies that $TU_tT^{-1} = U_{-t}$. This is a standard expression of the statement that time reversal is a dynamical symmetry.

Indeed, Definition 4.2 of dynamical symmetries helps explain a well-known property of dynamical symmetries. The solutions of a dynamical equation are associated with a representation $\varphi : t \mapsto \varphi_t$ of time translations: we have agreed that this is what it means for an equation to be 'dynamical'. In particular, any solution $\psi(t)$ with initial condition $\psi(0) = \psi_0$ can be written

$$\psi(t) := \varphi_t \psi_0, \tag{4.2}$$

for all time translations $t$. Let $s \in G$ be a dynamical symmetry, and let $S := \varphi_s$ be its representative on state space. Then, $sts^{-1} = t' \in \mathbb{T}$ implies that $S\varphi_t S^{-1} = \varphi_{t'}$ is a time translation, or equivalently $S^{-1}\varphi_{t'}S = \varphi_t$. Substituting this into Eq. (4.2) gives $\psi(t) = S^{-1}\varphi_{t'}S\psi_0$, which is equivalent to

$$S\psi(t) = \varphi_{t'}S\psi_0. \tag{4.3}$$

12 For a discussion of why unitaries and antiunitaries are the automorphisms of this state space, see Section 3.4.2.

https://doi.org/10.1017/9781009122139.005 Published online by Cambridge University Press

This is the familiar characterisation of dynamical symmetries expressed at the outset of this chapter – that whenever $S$ is a dynamical symmetry, if a curve $\psi(t)$ is a solution to the dynamical equation (Eq. 4.2), then the symmetry-transformed curve $S\psi(t)$ is also a solution (Eq. 4.3). Through an application of the Representation View and our definitions above, we can now offer an explanation of this statement: it follows whenever $S$ represents a symmetry of time translations in spacetime.

### 4.1.3 Time Reversal Symmetry

The discussion above allows us to classify temporal and dynamical symmetries into two basic types. When the time translation group is $\mathbb{T} = (\mathbb{R}, +)$, as in a typical frame of reference, there are only two automorphisms: the identity and the time reversal automorphism $t \mapsto -t$. This was proved in Proposition 2.1. Moreover, the definition of a temporal symmetry implies that it is an automorphism of $\mathbb{T}$. As a result, there are only two kinds of temporal symmetry $t \mapsto t' = sts^{-1}$: the (trivial) identity automorphism $sts^{-1} = t$ and the time reversal automorphism $sts^{-1} = -t$.

We can pull this classification down into a representation $\varphi : G \to \mathrm{Aut}(M)$ of a dynamical symmetry group as well. Defining $S := \varphi_s$, we have from the fact that a representation is a homomorphism that either $S\varphi_t S^{-1} = \varphi_t$ or else $S\varphi_t S^{-1} = \varphi_{-t}$. Thus, a dynamical symmetry preserves solutions in the sense of Eq. (4.3) in one of two ways:

$$S\psi(t) = \varphi_t S\psi, \qquad \text{or} \qquad S\psi(t) = \varphi_{-t} S\psi. \tag{4.4}$$

In the first case, the symmetry 'commutes' across time translations; in the second, it reverses time. *Time reversal invariance* is nothing more than an example of a symmetry of the second kind. We have seen an example of this at the end of Section 4.1.2.

Recall now from Chapter 3 that the time reversal transformation $T$ on state space is *by definition* the representative of $T := \varphi_\tau$ of a temporal reflection $\tau t \tau^{-1} = -t$, and which thus satisfies $T\varphi_t T^{-1} = \varphi_{-t}$. This means that a representation in which the time reversal element exists is, by definition, one that is time reversal invariant! Be careful though: this does not imply that every dynamical theory is time reversal invariant. Although we may extend time translations to a larger group including time reversal, there is no guarantee that a representation of that larger group exists.

As we will see in the next section, the non-existence of such a representation is precisely how the failure of a symmetry is expressed. This illustrates how deeply a symmetry may be encoded in a dynamical system:

the moment we describe time translations using the group $(\mathbb{R}, +)$, time reversal is a temporal symmetry; and, the moment we have a representation of time reversal on state space, time reversal is a dynamical symmetry. In other words, the meaning of the transformation arises fundamentally from its description as a symmetry. This is not a special or strange result about the meaning of time reversal but a standard, well-motivated feature of the Representation View of symmetry. We discuss it in more detail next.

## 4.2 Invariance and Non-invariance

### *4.2.1 The Symmetry Existence Criterion*

The Representation View has novel consequences for the interpretation of dynamical symmetry and asymmetry, which are worth stating clearly. Namely, a representation of a symmetry group plays at least two different semantic roles. The first involves reference: a representation of a spacetime symmetry group allows us to give meaning to those symmetries on state space and refer to them as 'spatial translations', or as 'time translations' or 'time reversal'. The second involves a symmetry property: a dynamical symmetry group is by definition a representation of a temporal symmetry group. So, in the special case of temporal symmetries, a representation plays two roles: it determines the very meaning of those transformations on state space, in addition to determining that they are dynamical symmetries. This second role can be summarised in the following:

(SEC) *Symmetry Existence Criterion.* If a representation of a temporal symmetry group exists, then it is a dynamical symmetry group ('invariance'); if no representation of a group exists, then it is not a dynamical symmetry group ('non-invariance').

A curious consequence is that if a spacetime symmetry group $G$ does happen to be a temporal symmetry group, then the failure of a representation to exist has consequences for reference: not only is there no symmetry, but there is no state space transformation that can meaningfully be interpreted as that spacetime symmetry. When a dynamical symmetry fails, it fails altogether to appear in the state space.

For example, rigid spatial translations often form a temporal symmetry group, as they do in the Poincaré group. This means that our statements about state space automorphisms, like 'the unitary operator $R$ representing rotations', are always made under the implicit assumption that a representation exists. If rotation is not a dynamical symmetry, in that a representation

https://doi.org/10.1017/9781009122139.005 Published online by Cambridge University Press

of rotations does not exist, then we have no good reason to view any unitary $R$ as referring to a rotation. In such cases, statements about 'the unitary operator $R$ representing rotations' are not even wrong: they are meaningless.[13]

Because of the dual role that a representation plays in characterising both symmetry and reference, the Symmetry Existence Criterion requires us to reinterpret what it means to *fail* to be a dynamical symmetry. After all, it is common practice to check that a given temporal symmetry is *not* a symmetry of a dynamical system. For example, in the spherically symmetric potential (recall Figure 4.2), a representation of time translations has the property that spatial translations are not generally dynamical symmetries. We often check this by explicitly verifying that an operator resembling spatial translations (such as the unitary $U_s = e^{iaP}$ in quantum theory) fails to preserve the solution space.[14] So, what are we saying when we do this?

When a state space representation of a spacetime symmetry exists, there are often arguments that it is uniquely defined: for example, in Chapter 3, considerations of the Galilei or Poincaré groups allowed us to show that if a representation of time reversal exists, then it is unique. Continuing the language of Section 3.4.3, let me refer to this unique definition as the 'appropriate' representation of a spacetime symmetry. What this means is:

*If* a representation of a temporal symmetry group exists in a dynamical system, *then* it has a particular canonical form that preserves the solution space.

The equivalent contrapositive statement is that: *if* the canonical form of a transformation like rotation or time reversal fails to preserve the solution space, *then* an adequate representation does not exist, and so the dynamical symmetry fails. The difference is subtle: if rotations are not dynamical symmetries, then it is the 'canonical form' of the rotations – but not a representation of them – that fails to preserve the solution space.

As a special case, we have seen above that the very existence of a representation $T$ of the time reversal group element $\tau$ encodes that time reversal is a dynamical symmetry. But, this does not mean that we have somehow illicitly assumed the existence of a symmetry that might not be found in nature. Such constructions should be viewed as conditional statements: that *if* time reversal is a dynamical symmetry *then* its representative $T$ must have a certain canonical form.

[13] There is more to say about how this can happen, of course, which will lead us into a discussion of symmetry breaking and symmetry violation: I will return to this in Section 4.5.

[14] Example: the solutions to the Schrödinger equation with spherically symmetric Hamiltonian $H = \frac{1}{2m}P^2 + 1/Q$ are not preserved by the spatial translation operator $U_s = e^{isP}$ when $s \neq 0$.

Thus, when I derive the canonical form of the time reversal operator using an assumption that amounts to time reversal symmetry, Callender (Forthcoming) may rest assured regarding his worry that "this assumption is a large one, for it's up in the air whether quantum mechanics is time reversal invariant". Time reversal invariance is not assumed in this derivation, for there is no guarantee that the time reversal operator exists. If time reversal invariance fails, then the operator $T$ will fail to exist as well. Equivalently, if a representation $T$ of time reversal $\tau$ does exist, then time reversal invariance is sure to hold. This is why we are allowed to assume time reversal invariance in most derivations of the meaning of time reversal, including the original derivation of Wigner (1931, §20).

### *4.2.2 Illustration in the History of Superselection*

Although the Symmetry Existence Criterion is not often made explicit in informal discussions of symmetry, it is implicitly adopted by Eugene Wigner and other founders of the Representation View. An illustration of this can be found in one of the interesting physical applications of time reversal. Namely, time reversal appears in the surprising discovery that some pairs of pure quantum states never compose to form a pure state: all of their non-trivial superpositions are mixed states.[15] This phenomenon, known as *superselection*, was first discovered by Wick, Wightman, and Wigner (1952), who derived the so-called fermion or univalence superselection rule: that there is no pure superposition of a pure boson state (of integer-multiple-of-$\hbar$ angular momentum) with a pure fermion state (of half-integer-multiple-of-$\hbar$ angular momentum).[16]

Remarkably, Wick, Wightman, and Wigner made use of the time reversal operator in their argument. So, following the great shock of 1964 when CP symmetry and time reversal symmetry were shown to be experimentally violated, Hegerfeldt, Kraus, and Wigner (1968) wrote a follow-up article entitled "Proof of the Fermion Superselection Rule without the Assumption of Time-Reversal Invariance", which derived the same superselection rule

[15] A *state* on a (unital $C^*$) algebra of observables as a positive linear unit-preserving functional $\omega : \mathcal{A} \to \mathbb{C}$. A *mixed* state is one such that $\omega = \lambda\omega_1 + (1-\lambda)\omega_2$ for some $\lambda \in (0, 1)$ and some states $\omega_1 \neq \omega_2$; otherwise $\omega$ is called *pure*.

[16] According to Wightman (1995, p.752), superselection originated in Wigner's suggestion in the 1940s that some self-adjoint operators are not observables. But as Earman (2008, p.379) points out, all the considerations needed for the fermion superselection rule were already in the Wigner (1931, §20) analysis of time reversal. See in particular Proposition 4.1. See Ruetsche (2004) for a discussion of the subtleties of interpreting mixed states and Earman (2008) for a philosophical appraisal of superselection.

using the spatial rotation operator instead of time reversal. The later authors explained their motivation:

> The fermion or univalence superselection rule, which separates states with integer and half-integer angular momentum, was originally proved under the assumption of time-inversion invariance. Recent experiments on $CP$ violation, combined with the $CPT$ theorem, now seem to question $T$ as a rigorous symmetry. Another proof of the fermion superselection rule without the assumption of $T$ invariance is thus desirable. (Hegerfeldt, Kraus, and Wigner 1968, p.2029)

The very existence of this later article would be puzzling if 'invariance' were taken to mean just 'preservation of the solution space' under an operator with the canonical form of time reversal, since, *the original derivation of fermion superselection did not assume a solution space is preserved.* In the original article, Wick, Wightman, and Wigner (1952, p.103) make no explicit mention of an equation of motion or solution space but simply postulate that "a time inversion operator ... exists" which has the property that $T^2 = -\mathbb{1}$ for a one-fermion system, while $T^2 = \mathbb{1}$ for a one-boson system (see Section 3.4.4). From this one can show that every non-trivial superposition of a boson state $\phi^+$ and a fermion state $\phi^-$ is 'mixed': if $\psi = \alpha\phi^+ + \beta\phi^-$, then $\langle\psi, A\psi\rangle = \lambda\langle\phi^+ A\phi^+\rangle + (1-\lambda)\langle\phi^-, A\phi^-\rangle$ for some $\lambda \in (0, 1)$ and for all bounded operators $A \in B(\mathcal{H})$ on the Hilbert space. For completeness, here is a reconstruction of the elegant original argument of Wick, Wightman, and Wigner (1952), in which only the existence of a time reversal operator is assumed.

**Proposition 4.1** (Wick, Wightman, Wigner 1952) *Let $\mathcal{H} = \mathcal{H}^+ \oplus \mathcal{H}^-$ be a direct sum of Hilbert spaces. If there exists a unitary or antiunitary $T : \mathcal{H} \to \mathcal{H}$ such that*

$$T^2\phi^+ = \phi^+ \text{ for all } \phi^+ \in \mathcal{H}^+ \qquad T^2\phi^- = -\phi^- \text{ for all } \phi^- \in \mathcal{H}^-, \tag{4.5}$$

*then every superposition $\psi = \alpha\phi^+ + \beta\phi^-$ with $|\alpha|^2 + |\beta|^2 = 1$ and $\alpha, \beta \neq 0$ is a mixed state.*

*Proof* Our assumptions imply that $T^2$ is unitary and commutes with all $A \in B(\mathcal{H})$. Thus,

$$\begin{aligned}\langle\phi^+, A\phi^-\rangle &= \langle T^2\phi^+, T^2A\phi^-\rangle = \langle T^2\phi^+, AT^2\phi^-\rangle = \langle\phi^+, A(-\phi^-)\rangle \\ &= -\langle\phi^+, A\phi^-\rangle.\end{aligned} \tag{4.6}$$

Hence, $\langle\phi^+, A\phi^-\rangle = 0$. A symmetric argument shows $\langle\phi^-, A\phi^+\rangle = 0$. So, for any superposition $\psi = \alpha\phi^+ + \beta\phi^-$ such that $|\alpha|^2 + |\beta|^2 = 1$,

$$\begin{aligned} \langle \psi, A\psi \rangle &= \langle \alpha\phi^+ + \beta\phi^-, A(\alpha\phi^+ + \beta\phi^-)\rangle, \\ &= |\alpha|^2 \langle \phi^+, A\phi^+ \rangle + |\beta|^2 \langle \phi^-, A\phi^- \rangle \end{aligned} \tag{4.7}$$

for all $A \in B(\mathcal{H})$. Thus, $\psi$ is mixed for all non-trivial ($\alpha, \beta \neq 0$) superpositions. ■

In this original theorem, there is no mention of an equation of motion, and so there is no solution space to be preserved. Why then does the later article by Hegerfeldt, Kraus, and Wigner (1968) state that the failure of time reversal symmetry invalidates it? I hope you'll agree that these authors, if anyone, are unlikely to be confused about the matter: Wigner in particular is the father of both time reversal and superselection.

The answer, I believe, is that they are implicitly adopting the Representation View on symmetry and are therefore committed to the Symmetry Existence Criterion. On this perspective, although we might not have said anything about the dynamics, the existence of a meaningful representation of the time reversal operator $T$ implicitly assumes that there is one, and so by the Symmetry Existence Criterion it is a dynamical symmetry. Equivalently, the discovery that time reversal invariance fails implies that a representation of the time reversal operator does not exist! This explains how the 1964 discovery of time reversal symmetry violation might have led Wigner and others to rethink the fermion superselection rule.

As the astute reader may note, the assumptions of the operator $T$ in Proposition 4.1 are also satisfied by the CPT operator $\Theta$ constructed by Jost (1957, 1965), which is guaranteed to exist in a very general sense by the CPT theorem.[17] So, although Wick, Wightman, and Wigner (1952) would not have known the possibility of replacing $T$ with $\Theta$, the analysis of rotations by Hegerfeldt, Kraus, and Wigner (1968) is not needed to restore the fermion superselection rule: the CPT operator $\Theta$ could have been used instead. However, the later result derived from rotational symmetry is still an independently interesting study of rotations and superselection. And, this seems to have been the main motivation for the authors of the second article.[18]

17 See Chapter 8.

18 In personal correspondence, Hegerfeldt tells me that in 1967, he and Kraus had written a paper on why a spin system's change of phase under a global rotation through $2\pi$ is unobservable, but which was rejected by a referee. Wigner was the leading expert on symmetry at the time. Hegerfeldt recalls, "we had never met him, but, young as we were, sent him a copy of our paper and of the referee report and asked for his advice". Wigner wrote back with extensive advice, which led the paper to eventually be published, and their correspondence afterwards led to the publication of Hegerfeldt, Kraus, and Wigner (1968). Charmingly, years later in 1980, Hegerfeldt reports that he

### 4.3 Direct Empirical Significance

With a clear expression of dynamical symmetry and time reversal symmetry in hand, it is possible to add some clarity to the three philosophical discussions of symmetry introduced at the outset of this chapter and to illustrate the role that time reversal plays. We begin in this section with the question of how we come to know about a dynamical symmetry and a review of the literature on direct empirical significance; I will turn to the second and third discussions in Sections 4.4 and 4.5.

Direct empirical significance has been explored as a way of clarifying the epistemology of symmetry: by interpreting symmetries as transformations of a subsystem with respect to its environment. For example, Galileo (1632) famously noticed that experiments done in an isolated cabin below deck on a ship proceed in the same way whether the ship is at rest or in uniform motion. This motivates the Principle of Relativity, that the laws of physics are the same in every inertial reference frame. However, Galileo did not arrive at his conclusion through the impossible task of boosting the universe. He arrived at it by observing a ship, which was in motion with respect to the sea. By breaking the boost symmetry with a comparison to the sea – for example, by looking through the porthole of the cabin – we provide a means of knowing that the two states of motion are actually distinct.

This observation led Kosso (2000) and Brading and Brown (2004) to postulate that two components are needed in order to explain how we know about symmetries: (1) a subsystem (the ship's cabin) on which a transformation is a dynamical symmetry; and (2) an environment (the sea) with respect to which the transformation is not a symmetry. Although two symmetry-related descriptions may be empirically indistinguishable within the subsystem, our ability to distinguish them in relation to the environment explains how we come to know about the symmetry empirically. The existence of these two components is called the *direct empirical significance* of a symmetry or of a symmetry group.

Direct empirical significance has been discussed for rotations, spatial translations, and boosts, as well as for tricky cases like gauge symmetries.[19] But, I am not aware of any comment about the empirical significance of time translations, let alone discrete symmetries like time reversal. The Representation View offers a helpful perspective on both. I will argue here

met a physicist on a bus going to a conference in Mexico, and that this physicist revealed that he "had been particularly happy about our paper because he had been the referee on that paper, had rejected it, but had been overruled". (Hegerfeldt, personal correspondence)

[19] Compare Gomes (2019, 2021a,b), Greaves and Wallace (2014), Healey (2007, Chapters 6–7), and Teh (2016).

that there are situations in which both time translation and time reversal have direct empirical significance. However, for discrete transformations like time reversal, the empirical significance piggy-backs on that of time translation symmetry. I will use the Fregean phrase *higher-order* to refer to this kind of direct empirical significance, for reasons that I will soon explain.

We begin by adopting the Representation View, in particular on what it means to have a state space representation of time translations.[20] Thus, time translation symmetry is what captures the repeatability of scientific experiments in time. If a description is empirically supported in the lab on Tuesday, then further confirmation may be found on Wednesday by repeating the same experiment. Formally speaking, this is expressed by the postulate that the group of time translations $\mathbb{T}$ can be represented among the automorphisms of a state space $\varphi : \mathbb{T} \to \mathrm{Aut}(M)$. I will adopt the simplest case of the time translation group, $\mathbb{T} = (\mathbb{R}, +)$.

In order to know empirically that an experiment has been repeated at different times, we need a way to compare those times, just as with Galileo's ship. One way to do this is to introduce a 'clock' into the environment. A clock can be a mechanical device that keeps time in the familiar way, or the changing brain state or age of an experimenter that perceives time's passage. This is not just an informal notion: there are formal accounts of clocks and general time-keeping environments, known as 'time observables', which allow the precise modelling of this situation. A time observable is one that keeps track of the parameter $t$ associated with time translations. For example, in quantum systems, a time observable $\mathcal{T}$ is one that tracks the unitary representation of time translations $U_t = e^{-itH}$, in that for any curve $\psi(t) = U_t\psi_0$ with initial state $\psi_0$ and initial expected time $t_0 = \langle\psi_0, \mathcal{T}\psi_0\rangle$, we have that $\langle\psi(t), \mathcal{T}\psi(t)\rangle = t_0 + t$. There is a great deal of debate over the nature and existence of time observables, which intersects the subtle question of what counts as an 'observable'.[21] However, it is enough for my purposes that in most cases, a time observable system is well-defined.

Given an environment with a clock system in this sense, we can give direct empirical significance to time translations in a way that is analogous to Galileo's ship: instead of comparing the ship to the sea, we compare it to the readings of a clock, as in Figure 4.4. When the initial conditions of the (ship) subsystem are restored from one day to the next, we can confirm that the

[20] See Section 3.1.2.

[21] It is well-known that such a time observable $T$ is not generally self-adjoint; however, it can still be modelled as a maximal symmetric or POVM observable. See Busch, Grabowski, and Lahti (1994), Galapon (2009), or Pashby (2014) for an overview, Hegerfeldt and Muga (2010) for a particularly general result, and Roberts (2014, 2018) for my own perspective.

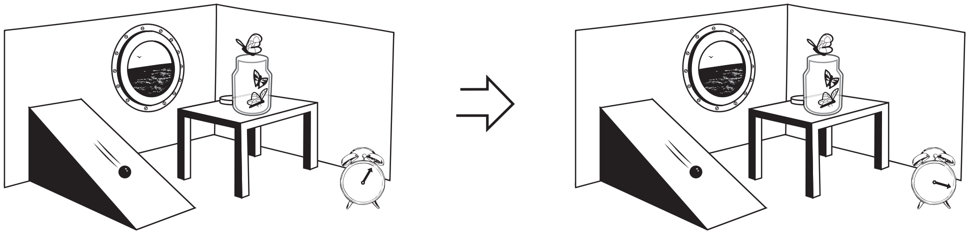

Figure 4.4 A clock system defines the direct empirical significance of time translations.

same behaviour ensues. This can be modelled using a representation of time translations among the automorphisms of the subsystem state space and thus justifies referring to them as symmetries. But, the clock subsystem is not restored: this captures the sense in which an observer can compare the two experiments as distinct. Note that the problem of defining the inverse time translations is not necessarily intractable: they can be defined operationally using a clock running in the reverse direction.[22] Thus, in the same way that the sea environment gives direct empirical significance to boosts, a clock environment gives direct empirical significance to time translations.

The direct empirical significance of time reversal is subtly different from this. I have argued that the most natural way to understand discrete symmetries like time reversal is as an automorphism of a continuous group of symmetries (Section 2.6). For example, time reversal is an automorphism of time translation symmetry, spatial reversal is an automorphism of spatial translation symmetry, and so on. In other words, a discrete symmetry is a 'symmetry of a symmetry'. The time translations $\mathbb{T}$ describe a structural property of time; but, since time reversal symmetry describes a structural property of $\mathbb{T}$, it is really a 'property of a property' of time. This is what I mean when I say that it is 'higher-order'.

Being higher-order does not make time reversal any less directly empirically significant. When it exists, the time reversal automorphism is a property of the group of time translations, which themselves have direct empirical significance. For example, if we are able to grasp that the group structure of time translations is $\mathbb{T} = (\mathbb{R}, +)$ in some context, then we can immediately deduce that it has a time reversal automorphism $t \mapsto \tau t \tau = -t$. So, if we have direct empirical access to time translations, then our empirical access to time reversal symmetry is a 'free lunch'. This higher-order property is analogous

[22] See Section 2.2, Figure 2.3. From the perspective of time observables, this corresponds to the fact that for every time observable $\mathcal{T}$, there is a 'reversed' time observable $-\mathcal{T}$.

to Frege's example[23] that 'being unique' is a higher-order property of 'being a moon of the Earth', which is itself directly empirically accessible through the study of 'the Moon'. And, of course, a similar interpretation is available for other discrete symmetries like parity and parity–time reversal as well.

The argument I am giving here can be summarised as follows.

1. If a group of symmetries has direct empirical significance, then all of its properties have direct empirical significance too.
2. The group of time translations has direct empirical significance, and one of its properties is time reversal symmetry.
3. Therefore, *time reversal symmetry has direct empirical significance too*.

The first premise is a condition of adequacy for a reasonable account of direct empirical significance. The second premise follows from the arguments I have given above: time translation symmetries are directly empirically accessible through the study of a subsystem with respect to a clock environment; and, time reversal is a property of the group of time translations, since it is by definition one of its automorphisms. It follows that time reversal has direct empirical significance. Again, a similar argument can be made about parity and parity–time reversal, too.

Note that I am *not* assuming that a symmetry with 'direct empirical significance' can necessarily be operationally implemented. In most contexts, there is no way to continuously transform a system to its time reverse: in the full Poincaré group, the time reversal operator $\tau$ is not continuously connected to the identity, which makes such a continuous transformation impossible. But, other discrete transformations, like the parity–time transformation, are continuously connected to the identity in the universal covering group of the Poincaré group (discussed in Chapter 8). Moreover, I find it helpful to separate the question of whether we can 'operationally implement' a symmetry from the question of whether we have direct empirical access to it. Direct empirical significance is about whether we can know about a symmetry through the comparison of a subsystem to its environment. In the case of time reversal, we can.

## 4.4 The Symmetry-to-Reality Inference

Leibniz wrote in his third letter to Clarke that the universe would be left unchanged by a transformation that involves only "changing East into West" (Leibniz and Clarke 2000, p.14). This famously inspired Earman and Norton

[23] Cf. Frege (1884, §53).

https://doi.org/10.1017/9781009122139.005 Published online by Cambridge University Press

(1987, p.522) to define 'Leibniz Equivalence' in general relativity to mean that, "[d]iffeomorphic models represent the same physical situation", where 'diffeomorphic models' are those related by a symmetry. The *symmetry-to-reality inference* is a generalisation of these principles, which states: if two descriptions are related by a symmetry, then they refer to the same physical situation. The term was coined by Dasgupta (2016, p.840), who formulates it as the statement, "if a putative feature is variant in laws that we have reason to think are true and complete, then this is some reason to think that the feature is not real". More standard usage reformulates this in an equivalent form: if a feature of a theory correctly describes reality, then it is invariant under the symmetries of that theory.

Although some commentators have suggested that the symmetry-to-reality inference is always or nearly-always justified, most argue that it is not.[24] In this section I would like to add two additional cautionary remarks about situations in which the symmetry-to-reality inference is either ill-posed or invalid. It is ill-posed in cases where a physical description is mentioned without specifying its reference. And, when the reference of a description is well-specified, there is still a sense in which it is invalid, just as it is invalid to infer that the Klein four-group has only one subgroup of order two. Time reversal will provide one counterexample to the claim.

The first cautionary remark stems from a more general comment about language-to-reality inference, in the influential essay on "Sense and Reference" by Frege (1892). Frege described a language community's common understanding of how to use a proper name as the 'sense' of the name and the object in reality that the name refers to as its 'reference' or 'meaning' (*Bedeutung*). He gave a famous example of this, 'the morning star' and 'the evening star', which he took to have different senses but the same reference: the planet Venus. His cautionary remark about sense and reference was the following.

It may perhaps be granted that every grammatically well-formed expression representing a proper name always has a sense. But this is not to say that to the sense there also corresponds a reference. The words 'the celestial body most distant from the Earth' have a sense, but it is very doubtful if they also have a reference. The expression 'the least rapidly convergent series' has a sense but demonstrably has

[24] Something like this inference appears in Baker (2010), Fletcher (2018), and Weatherall (2018); see also Baker (2020, "What are symmetries?" In: Unpublished Manuscript, http://philsci-archive.pitt.edu/18524/, where it is true by definition. In contrast, Belot (2013, p.330) gives a number of examples showing the symmetry-to-reality inference (which he calls 'D2') is "false if understood as a thesis concerning classical symmetries of differential equations". Caution is similarly advised by Butterfield (2021), De Haro and Butterfield (2019), Møller-Nielsen (2017), Read and Møller-Nielsen (2020), and myself (Roberts 2020).

no reference, since for every given convergent series, another convergent, but less rapidly convergent, series can be found. In grasping a sense, one is not certainly assured of a reference. (Frege 1892, p.28)

The same is often true of models in physics. Suppose we are speaking not just of proper names but about whole sentences, and indeed about whole mathematical structures containing sentences, like a model of a quantum system. A model in physics often has sense without reference, which makes any symmetry-to-reality inference impossible. To take just one example: physics students study all manners of facts about the simple harmonic oscillator as a dynamical system – its symmetries, energy levels, basis representations, perturbative approximations, and many other things – without saying anything about which physical system is being referred to. Indeed, part of the power of the harmonic oscillator is that it can refer to countless different things: it is the first-order approximation of any system described by a locally-characterised (meaning, analytic) force.[25] In the use-mention jargon of Quine (1940, §4), the textbook exercises about the harmonic oscillator might thus be viewed as 'mentioning' their use in situations where they have reference, rather than 'using' them as such. De Haro and Butterfield (2018) draw a similar conclusion, pointing out that symmetries (and especially dualities) are often defined in the context of uninterpreted 'bare theories'.

Thus, when the reference of a dynamical system is not fixed, there may be no answer to the question of whether symmetry-related descriptions refer to the same thing: in such cases, the question is not well-posed. So, let us suppose the reference of a given physical description *is* well-posed. I claim that the symmetry-to-reality inference is still not generally valid. This leads to the second remark.

Dasgupta (2016, p.840) formulates the symmetry-to-reality inference as applying to any "putative feature" of reality as it appears in a theory. At this level of generality, there are simple counterexamples to the inference. For example, consider the description of a free Newtonian particle. To avoid formulating an ill-posed question, suppose that each possible solution in state space is identified with some possible trajectory for a physical particle in the absence of forces, each with a different (constant) velocity, as in Figure 4.5. Suppose also that a 'putative feature' of reality is associated with

[25] Recall that a harmonic oscillator is a dynamical system with a quadratic potential $U = ax^2$ arising from a linear force $F(x) = -\nabla U = -2ax$. Its widespread usefulness is explained by the fact that if $F'(x)$ is any other force that is 'locally characterised' in the sense of being an analytic function, then $F'(x)$ has a Taylor expansion and is therefore linear in its first-order approximation. In other words: the harmonic oscillator is a first-order approximation of every locally-characterised force.

https://doi.org/10.1017/9781009122139.005 Published online by Cambridge University Press

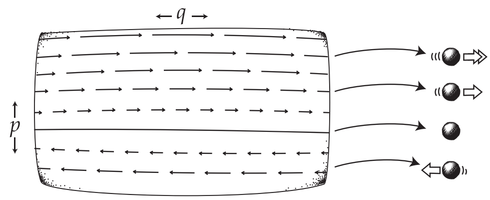


Figure 4.5 Choosing the 'reference' of a dynamical system by associating each curve with a possible motion.

one of the trajectories with non-zero velocity. Is it preserved by dynamical symmetries, as the symmetry-to-reality inference requires?

The answer is, 'no'. For example, a representation of parity, which transforms positions and velocities as $\pi : (x, \dot{x}) \mapsto (-x, -\dot{x})$, is a dynamical symmetry of this system.[26] But, it does not preserve each individual curve. A curve is rather transformed to one with opposite velocity, 'reflecting' the phase space diagram in Figure 4.5 about both axes. So, this 'putative feature' of reality is not preserved by a symmetry.

One might respond by observing that, by refining the symmetry-to-reality inference, it is possible to avoid this counterexample. Namely, instead of applying the inference to arbitrary 'putative features' of reality, like individual trajectories, suppose we apply it to the space of all trajectories as a whole. That is, consider the following 'global version' of the symmetry-to-reality inference: that *if two representations of time translations are related by a dynamical symmetry, then they refer to the same physical time translations.*

This formulation of the inference uses 'reference' in a way that would require further analysis, in order to say what it means for a group of time translations (or equivalently, all the possible trajectories at once) to 'refer' to a physical time translation. One might take inspiration from what Dewar (2019) calls *sophistication about symmetries* and view the collection of all symmetry-related trajectories as being what picks out a description of reality, rather than any single trajectory in the solution space.[27] As Dewar writes:

In a slogan, the idea is that we need not insist on finding a theory whose models are invariant under the application of the symmetry transformation, but can rest content with a theory whose models are isomorphic under that transformation. (Dewar 2019, p.298)

[26] As a quick check, write the solutions of a classical particle as $x(t) = \dot{x}_0 t + x_0$ for each initial state $(x_0, \dot{x}_0) \in \mathbb{R}^2 \times \mathbb{R}^2$. Then a parity-transformed trajectory has the form $x^{\pi}(t) = -\dot{x}_0 t - x_0$, which is again a solution, with initial state $(-x_0, -\dot{x}_0)$.
[27] For an evaluation of sophistication, see Martens and Read (2020).

This does allow one to avoid the counterexample of parity. For the free particle, parity preserves the representation of time translations as a whole, $\pi\varphi_t\pi^{-1} = \varphi_t$. Spatial translation, spatial rotation, and time translation symmetry do as well. So, whatever the group of time translations $t \mapsto \varphi_t$ refers to, it is preserved by all these symmetries.

Unfortunately, there is a counterexample to this 'global' version of the inference too, which is the case of time reversal symmetry. Recall from Section 4.1.3 that every dynamical symmetry either preserves each individual time translation, $\varphi_t \mapsto \varphi_t$, or reverses it, $\varphi_t \mapsto \varphi_{-t}$. Time reversal is an example of the second case: it does not generally preserve the group of time translations as a whole.

Of course, viewed as a group, the time-reversed representation $t \mapsto \varphi_{-t}$ is isomorphic to the original, since time reversal is, after all, an automorphism. But, this does not mean that the representations of that group are the same. Recall the example of the Klein four-group at the outset: there is just one group of order two but multiple representations of it amongst the subgroups of the Klein four-group. Similarly, when time reversal is a dynamical symmetry, we have two representations of the time translation group $(\mathbb{R}, +)$, given by $t \mapsto \varphi_t$ and $t \mapsto \varphi_{-t}$. Although they are isomorphic, it would be wrong to say that they are one and the same.

One could try to save the symmetry-to-reality inference with yet more sophistication about symmetries. Or, one might replace the time translation group $(\mathbb{R}, +)$ with the quotient group $(\mathbb{R}, +)/\mathbb{Z}_2$ of equivalence classes of time translations of the form $(t, -t)$. If time translations are representations of that group, then time reversal would be what Caulton (2015) calls an 'analytic symmetry' and could not refer to different physical situations. However, the choice of structure to describe time translations has physical content, and so one should be wary of what this commits us to. For example, by adopting the quotient group $(\mathbb{R}, +)/\mathbb{Z}_2$ for time translations, we can only describe physical systems that evolve forwards and backwards in time in exactly the same way. In most dynamical systems in physics, this does not describe any realistic physical system, suggesting the approach is on the wrong track.

## 4.5 The Spacetime–Dynamical Symmetry Relationship

Our last philosophical discussion is about whether every spacetime symmetry is a dynamical symmetry, and vice versa. The tight relationship between spacetime and dynamical symmetries played a pivotal role in the development of spacetime physics, at least since Minkowski (1908, §II) argued that the symmetries of Lorentz's dynamical theories of matter "force this changed conception of space and time upon us". Earman (1989, §3.4), following Stein (1977, p.6), gave an influential discussion of that relationship, as part of his approach to interpreting spacetime theories.[28] His proposal was formulated in terms of two postulates:

(SP1) A dynamical symmetry is also a spacetime symmetry. And
(SP2) A spacetime symmetry is also a dynamical symmetry.

Earman himself viewed these principles as "conditions of adequacy on theories of motion" (Earman 1989, p.46). But, an interpretation of spacetime due to Brown and Pooley (2006) and Brown (2005) included an explanation of why this might be so, by arguing that the ghostly concept of 'spacetime structure' actually just encodes (a philosopher might say 'supervenes on') facts about the dynamics of matter fields. Most commentators have emphasised that this is unlikely to be the entire story, with dynamical systems inevitably including some spacetime concepts in their formulation.[29] I agree: as I argued in Section 2.3, a dynamical system must at least admit a representation of time translations, in order to justify being called 'dynamical'. However, many go beyond viewing Earman's principles as adequacy conditions and treat them as truths. Myrvold (2019) even argues that they are analytic truths, or true by the very meaning of the words they contain:[30]

> Both connections between dynamical symmetries and spacetime symmetries are analytic. To say that, for test bodies, there is a real distinction between motion along geodesics of the metric and motion that is forced away from geodesics, is to simultaneously refer to a fact about the dynamics of these bodies and their spacetime environment. To say that the field equations do (or do not) require a flat background spacetime is to simultaneously refer to a fact about these field equations and their spacetime setting. (Myrvold 2019, p.142)

The Representation View leads to a different conclusion: both (SP1) and (SP2) can fail, in ways that are important for modern physics. So, these

[28] That context was the following. We first view a manifold for a spacetime theory as admitting two kinds of smooth (tensor or spinor) fields, the 'absolute' fields describing the structure of spacetime itself and the 'object' fields describing the material contents of spacetime, where the latter evolve according to some dynamical laws. A spacetime symmetry is then a transformation of the absolute fields, while a dynamical symmetry is a transformation of the object fields; both are defined to leave the theory's space of models invariant (Earman 1989, p.45). The approach to dynamical symmetries that I have set out in Section 4.1.2 is compatible with this.
[29] Cf. Lehmkuhl (2011), Maudlin (2012), Myrvold (2019), Norton (2008b), and Knox (2019).
[30] Similar conclusions are drawn by Acuña (2016, p.2) and Knox (2019, p.124).

principles are not analytic truths. However, the failure of (SP2) suggests that we should make some interesting adjustments to the mathematical structures we use to describe reality. In this case, Earman was correct: (SP2) is an adequacy principle for good scientific theorising. I will discuss each principle in turn.

### *4.5.1 Dynamical Symmetries that Are Not Spacetime Symmetries*

If a dynamical theory can be formulated on any interesting class of general relativistic spacetimes, then there is a sense in which (SP1) is false. To see this, let me first recall some definitions. For $G$ to be a spacetime symmetry group, its elements must preserve spacetime structure, the way that isometries preserve the structure of a Lorentzian manifold. For $G$ to be a temporal symmetry group, it must preserve the time translations $\mathbb{T}$ when it acts on them by conjugation (Definition 4.1). Finally, for $G$ to be a dynamical symmetry group on some state space, it must both be a temporal symmetry group and admit a representation on state space (Definition 4.2).

Consider now a generic relativistic spacetime with no isometries at all. This is the case for most spacetimes describing realistic gravitational phenomena. Then, of course, 'time translations' cannot be spacetime symmetries, no matter what we take them to be. But, the Representation View still requires that when we formulate a dynamical theory, we must adopt a representation of *some* time translations, although they might be local.[31] Otherwise, the theory is not deserving of the name 'dynamical'.

So, let $\mathbb{T}$ be a structure representing time translations, and suppose that it admits a representation on state space. To make the example simple, assume that the spacetime admits a smooth, complete, timelike vector field, describing the passage of time in some reference frame, and let $\mathbb{T} = (\mathbb{R}, +)$ be the group of time translations that flows along it. Then, $G = \mathbb{T}$ satisfies our definition of a dynamical symmetry group: it is a special case of a temporal symmetry group,[32] and it admits a representation. But, it is not a spacetime symmetry group. So, in any dynamical theory formulated on generic relativistic spacetimes (cf. Wald 1994), there are dynamical symmetries that are not spacetime symmetries.

[31] Non-orientable spacetimes by definition do not admit a smooth timelike vector field; in such cases, time translation may still be modelled in local neighbourhoods using a *local* Lie group, such as one associated with $(\mathbb{R}, +)$ in a neighbourhood of the identity; this was our strategy in local symplectic mechanics (Section 3.3.1).

[32] For $G$ to be a temporal symmetry group of $\mathbb{T}$, the elements of $G$ must preserve $\mathbb{T}$ when acting on it by conjugation, $gtg^{-1} \in \mathbb{T}$. This holds trivially when $G = \mathbb{T}$ by the group closure axiom.

### 4.5.2 Spacetime Symmetries that Are Not Dynamical Symmetries

The previous symmetry principle fails in the context of gravitation, where spacetimes can be expected to have no non-trivial symmetries. In contrast, (SP2) can fail even for local physics: there are reasonable models in which not every local spacetime symmetry is a dynamical symmetry! So, I do not see any sense in which this principle is analytic. However, when (SP2) fails, working to restore it usually leads to interesting interpretive adjustments to a theory. So, it is not unreasonable to side with Earman and view (SP2) as an adequacy principle for good physical theorising.

There are three ways that (SP2) can fail. Suppose that $G$ is a spacetime symmetry group, and recall that $G$ is a dynamical symmetry group only if it is both a temporal symmetry group and admits a representation on state space. This can fail to occur in three different ways. I will just summarise them here and then discuss their details in the next subsections:

1. *No temporal symmetries*. A spacetime symmetry group $G$ might not be a temporal symmetry group, meaning that it might not preserve the structure of time translations.
2. *No representation of temporal symmetries*. Even if $G$ is a temporal symmetry group, and even if a dynamical theory is given by a representation of time translations, there may be no non-trivial extension of that representation to all of $G$.
3. *No 'adequate' representation of temporal symmetries*. The group $G$ might be a dynamical symmetry group in some representation, but the representation still may not be 'appropriate' given the interpretive constraints associated with a state space, where I use the word 'appropriate' as in Section 4.2.1.

The first two possibilities are already recognised in some corners of the quantum field theory community, where the Representation View is perhaps most well-known. For example, Bogolubov, Logunov, Oksak, and Todorov (1990, p.375) write: "the term 'non-invariance' can mean that either the specified symmetry of the field algebra does not exist, or such a symmetry is not realizable (anti-)unitarily [in a representation]". The third possibility is not a 'strict' failure of (SP2) but is reason enough to doubt it. Let me explain each of these cases in more detail.

### 4.5.3 No Temporal Symmetries

A group of spacetime symmetries $G$ does not necessarily preserve the structure of time translations, in which case it is not a dynamical symmetry

https://doi.org/10.1017/9781009122139.005 Published online by Cambridge University Press

group. One has some freedom in choosing what we mean by time translations when we formulate a dynamical theory. I have already given one example of this in my discussion of the Lorentz boosts (Section 4.1.2): let $G$ be the group of symmetries of Minkowski spacetime, the Poincaré group. Suppose one chooses a smooth timelike vector field associated with a choice of reference frame and identifies a group of time translations $\mathbb{T} = (\mathbb{R}, +)$ associated with the flow along that reference frame. This is a typical choice in dynamical theories formulated in terms of the Euler–Lagrange equations, and in Hamiltonian's equations as well.

However, on this definition of time translations, Lorentz boosts are spacetime symmetries that are not temporal symmetries. In simple terms, this is because time translations are associated with a fixed reference frame, which is not preserved by velocity boosts. In the context of the Euler–Lagrange equations and Hamilton's equations, this arises in the well-known fact that boosts do not, strictly speaking, preserve the solutions to these dynamical equations. So, boosts are spacetime symmetries, but they are not dynamical symmetries because they fail to even be temporal symmetries.

An appropriate response to this issue is to expand the choice of time translation symmetries in a way that is more appropriate for the Poincaré group. As I showed in Section 4.1.2 (and especially Figure 4.3), choosing time translations to be the larger group that flows along any timelike geodesic restores boosts as temporal symmetries. This temporal symmetry group is preserved by all the spacetime symmetries, and so (SP2) is restored.

In summary, spacetime symmetries can fail to preserve our definition of time translations, in which case they are not dynamical symmetries. This failure of (SP2) is possible because of our freedom to choose what we mean by time translations. We can avoid that failure by introducing a further postulate: that our definition of time translations *must* be preserved by all the spacetime symmetries. However, there remain other ways that (SP1) can fail, which I turn to next.

### 4.5.4 No Representation of Temporal Symmetries

The second way that a spacetime symmetry can fail to be a dynamical symmetry is when no representation exists. Let us set aside the previous concern and assume in what follows that our group of spacetime symmetries preserves time translations, at least at the level of spacetime structure. That is, let us suppose that the spacetime symmetry group $G$ is a temporal symmetry group. The problem is that the structure of state space may still prohibit any non-trivial representation of the spacetime symmetry group, in which case it fails to be a dynamical symmetry group. I will begin with a simple example of this in the case of Hamiltonian mechanics and then argue that the failure of (SP2) is indeed essential to the phenomenon of spontaneous symmetry breaking.

*Failure in Hamiltonian Mechanics*

Let me begin with a simple example to illustrate. Suppose spacetime has the symmetries of the complete Galilei group, including time reversal. So, the time translations $\mathbb{T} = (\mathbb{R}, +)$ are spacetime symmetries as well. Suppose our dynamical theory is Hamiltonian mechanics, formulated on a symplectic manifold $(M, \omega)$ as in Section 3.3.1. Strictly speaking, the automorphisms of this state space are the symplectomorphisms, which are the transformations that preserve both $M$ and $\omega$: namely, they are diffeomorphisms, $\phi : M \to M$, that identically preserve the symplectic form, $\phi^*\omega = \omega$, where $\phi^*$ denotes the pull-back of a diffeomorphism $\phi$. So, a representation of time translations in this context is a homomorphism $\varphi : (\mathbb{R}, +) \to \mathrm{Aut}(M, \omega)$, where each time translation $\phi_t \in \mathrm{Aut}(M, \omega)$ is a symplectomorphism.

Now, the time reversal group element $\tau$ is a spacetime symmetry as well. But, if the Hamiltonian for these time translations (representing energy) is bounded from below but not from above, as is usually the case, then it turns out to be impossible to represent it as a symplectomorphism: Proposition 3.1 implies that it must be antisymplectic. The problem, roughly speaking, is that the structure of a symplectic manifold as a state space includes an orientation.[33] This makes it impossible to extend the representation of time translations to one that includes time reversal. So, strictly speaking, time reversal fails to be a dynamical symmetry of symplectic mechanics, even when it is a spacetime symmetry!

One might simply stop there and conclude that classical Hamiltonian mechanics is time reversal symmetry violating. However, a more natural response is to change what we mean by a state space automorphism: by extending the automorphisms to include both symplectic and *anti*-symplectic transformations, we can again view time reversal as a dynamical symmetry.[34] In this sense, the failure of (SP2) here is a flag that the structure of state space might need to be reexamined, which leads to an interesting revision of the theory.

This revision is not an analytic truth: it is not *meaningless* to assert that the automorphisms of a state space are symplectomorphisms and also that

[33] A symplectic form is commonly interpreted as the 'oriented area' of the parallelogram determined by two vectors at a point. See e.g. Arnol'd (1989, §7(C)).
[34] This is the approach taken in Section 3.3.

the symmetries of spacetime are given by the complete Galilei group. If anything, Earman's interpretation of (SP2) as a 'condition of adequacy' is most appropriate: (SP2) gives some indication that anti-symplectic transformations might be automorphisms of classical Hamiltonian mechanics as well.

*Failure for Spontaneous Symmetry Breaking*

Another more well-known example in which spacetime symmetries are not dynamical is the case of spontaneous symmetry breaking. Symmetry breaking is a cornerstone of quantum field theory, including the Higgs mechanism, for which degenerate ground states in an electroweak interaction are related by an $SU(2) \times U(1)$ symmetry. But, my point can be made in a simple toy example that is familiar to philosophers of physics: the half-infinite Ising lattice.[35] Let me sketch the structure of that system.

Let the spacetime symmetries be described by the Poincaré group, although any spacetime with rotational symmetries will do. To construct our state space, we will make use of the half-infinite *Ising lattice*, which is an infinite chain of spin-1/2 systems, each imagined to be spatially separated and labelled by the natural numbers, $n = 1, 2, 3, \ldots$, as in Figure 4.6. The algebra of observables describing this system is the Pauli algebra for a countably infinite number of spin systems: that is, for each spin-system $n = 1, 2, 3, \ldots$, we have three self-adjoint operators $\sigma_x^n, \sigma_y^n, \sigma_z^n$ satisfying the Pauli relations, and where any two observables associated with different spin systems on the chain are assumed to commute. The state space for this system is then a representation of the Pauli algebra amongst the operators on a Hilbert space $(\mathcal{A}, \mathcal{H})$. The Hilbert space $\mathcal{H}$ is required to be 'separable', meaning that its dimension is countable; this ensures that some of the basic interpretive structures of quantum theory can be applied.[36]

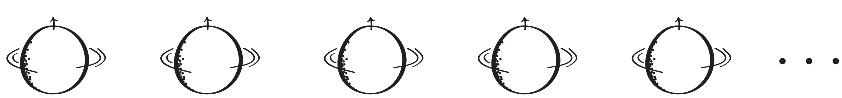

Figure 4.6 A half-infinite Ising lattice.

[35] This toy model has been the subject of much philosophical discussion: see Baker (2011), Baker and Halvorson (2013), Feintzeig (2015), Jacobs (Forthcoming), Kronz and Lupher (2005), Ruetsche (2006, 2011), and D. Wallace (2018, "Spontaneous symmetry breaking in finite quantum systems: A decoherent-histories approach", Unpublished manuscript, http://philsci-archive.pitt.edu/14983/).

[36] See Sewell (2002, §§2.3.1) for the construction of this representation and J. Earman (2020, "Quantum physics in non-separable Hilbert spaces", Unpublished manuscript, http://philsci-archive.pitt.edu/18363/) for a recent philosophical discussion of separability.

https://doi.org/10.1017/9781009122139.005 Published online by Cambridge University Press

One of the surprising but well-known features of the half-infinite Ising lattice is that, when we reverse two of the coordinate axes and build a new representation of the Pauli algebra, corresponding to a global rotation in space through the angle $\pi$, the result is not 'unitarily equivalent' to the original. This is what it means to say that spontaneous symmetry breaking occurs. More precisely, calling the first representation $(\mathcal{H}^+, \mathcal{A}^+)$ and the second representation $(\mathcal{H}^-, \mathcal{A}^-)$, unitary inequivalence means that there is no unitary map $R : \mathcal{H}^+ \to \mathcal{H}^-$ such that $A \mapsto RAR^{-1}$ is a bijection from $\mathcal{A}^+$ to $\mathcal{A}^-$. The structure of this state space prohibits unitarily equivalence.[37]

For our purposes, what is significant about this fact is that, if a representation of the spatial rotation group exists, then it *must* express unitary equivalence: if rotations were not unitary, then the representation would not consist in automorphisms of state space (see Section 3.4.2); and, if it did not relate spatially rotated descriptions, then it would not be homomorphic to the rotation group. So, the half-infinite Ising lattice is a system in which *no representation of the rotation group exists*. As a result, rotations cannot be dynamical symmetries, no matter what we take the dynamics to be, in spite of the fact that they are spacetime symmetries. Thus, we get another failure of (SP2).

A classic response to this situation is to argue that the appropriate state space for quantum field theory is fundamentally the abstract $C^*$ algebra defining the Pauli relations, rather than a Hilbert space representation of this algebra, as is suggested by Segal.[38]

> The proper sophistication, based on a mixture of operational and mathematical considerations, gives however a unique and transparent formulation within the framework of the phenomenology described; the canonical variables are fundamentally elements in an abstract algebra of observables, and it is only relative to a particular state of this algebra that they become operators in Hilbert space. (Segal 1959, p.343)

I will not take a position on this view. The lesson that I would like to draw is that, in the presence of spontaneous symmetry breaking, the status of (SP2)

[37] In this construction, that structure can be viewed as arising from the Hilbert space not having 'enough' basis vectors for an infinite system. One would like to view each spin site as having two orthogonal basis vectors representing 'spin-up' and a 'spin-down' with respect to some axis. But, with a countable set of spin sites, this would mean that there are $2^{\aleph_0}$ basis vectors, and the Hilbert space would not be separable. To avoid this, one can generate a basis set by beginning with the basis vector in which every spin site is 'up' in some direction and then introducing the countable set of basis vectors that are 'down' at only finitely many sites. In contrast, the basis set for a rotated representation, only a finite number of spin sites are 'up'. As a result, global polarisation becomes a representation-dependent property, and unitary equivalence fails. For details, see Sewell (2002, §2.2.3).

[38] See Feintzeig (2015, 2018) for a recent defence of this view.

can again be used to motivate rethinking the structure of state space. As Segal suggests, the choice is not an analytic truth but requires subtle philosophical as well as 'operational and mathematical' considerations. Again, (SP2) is not analytic but rather an interesting adequacy condition for theorising. Indeed, when we turn to the discussion of dynamical symmetry violation in Section 7.1, I will argue that with a bit of empiricism in mind, it can still be reasonably applied.

### 4.5.5 No Adequate Representation

The final example of how (SP2) can fail involves an issue that is commonly referred to as 'symmetry violation'. It might happen that we have a spacetime symmetry that preserves time translations, and even have a representation of that spacetime symmetry. But, our representation might not be 'appropriate' given the interpretive constraints of our theory.

To illustrate what it means to have an 'inappropriate' representation, recall the example from Section 3.2.4 of time reversal for the free particle. For this system, an 'appropriate' representation of the time reversal group element $\tau$ is given by $T(x,\dot{x}) := (x,-\dot{x})$. But, we can also construct an 'inappropriate' transformation $\tilde{T}$ defined by $\tilde{T}(x,\dot{x}) := (-x,\dot{x})$. Our previous calculation revealed[39] that $\tilde{T}$ provides a representation of time reversal. As a result, $\tilde{T}$ reverses each trajectory in the same way that time reversal does, $x(t) \mapsto x(-t)$. But, most would consider this $\tilde{T}$ to be an appropriate representation of space-and-time reversal, and not of time reversal alone. This is what I mean by the statement that $\tilde{T}$ is an inappropriate representative of time reversal.

Potential trouble for (SP2) arises when we have *some* representation of a temporal symmetry but no 'appropriate' representation of it. For example, in the context of quantum theory, it is possible to show that when energy is half-bounded, there is always *some* representation of time reversal: this was the first part of Proposition 3.4 in Chapter 3. In spite of this, it is not generally possible to find a time reversal operator that reverses 'time and only time'. That is what happens when time reversal symmetry violation occurs, and it is known to occur in the theory of electroweak interactions. In such cases, we have no appropriate representation of time reversal, even when time reversal is a symmetry of the background spacetime.

There is a reasonable response to this too. Instead of rejecting (SP2), we might take the discovery of time reversal symmetry violation to imply that

[39] See especially Eq. (3.10) of Section 3.2.4.

we should revise our spacetime structure: since time reversal is a symmetry of Minkowski spacetime, then perhaps Minkowski spacetime by itself does not accurately describe local spacetime structure.

I agree with this conclusion. However, there remains a gap in the argument, which is to establish why time reversal symmetry violation requires a revision of spacetime structure and not just the rejection of (SP2). Fortunately, this is not a very large gap: I will argue in Section 7.1 that just a little bit of empiricism is enough to choose the former. But, I reserve this argument for Chapter 7, where we discuss the implications of dynamical symmetry violation for the arrow of time in more detail.

## 4.6 Summary

In this chapter I have set out some implications of the Representation View for the philosophy of symmetry. We found that the concept of a symmetry splits into two components: temporal symmetries on spacetime and dynamical symmetries on state space. A representation then plays the dual role by giving meaning to a transformation on the state space of a dynamical theory and also by expressing that it is a symmetry of that theory. This gave rise to what I called the Symmetry Existence Criterion and in particular the fact that if a temporal symmetry fails to be a dynamical symmetry, then there is no representation of that symmetry and so no transformation on state space that can be meaningfully said to correspond to it.

The Representation View also suggested some revisions to three general debates on the philosophy of symmetry: direct empirical significance, the symmetry-to-reality inference, and the relation between spacetime and dynamical symmetries. In the next chapter, I will turn to a fourth: the infamous arrow of time.