Symmetries in Physical Reasoning

Shamik Dasgupta

Draft of February 2010

The concept of symmetry plays a central role in physical reasoning, both by philosophers of physics and physicists themselves. That much is not news. However, a puzzle emerges when these roles are put under scrutiny. For once the inferences involving symmetry are set out it is natural to ask whether they are justified, but when one examines the literature in which these inferences are made one finds either that no justifications are given, or that the justifications on offer are woefully inadequate. Of course, the natural response is to justify the inferences oneself. But when looking for a justification it is natural to clarify the content of the inferences by seeing how 'symmetry' is defined, and it turns out that on *any* definition in the literature the inferences are at best unjustifiable, at worst plain mistaken! So the puzzle is that while these inferences involving symmetry are central to physical reasoning, it would appear on further examination that they are entirely ungrounded.

One reaction to this puzzle would be to proclaim that a vast swarth of physical reasoning is based upon a mistake. But that is not my reaction. I think that the inferences involving symmetry *can* be justified, so my aim in this paper is to offer a "rational reconstruction" of this fragment of physical reasoning. The reconstruction will consist of two stages. In the first stage I will argue that one kind of inference involving symmetry is epistemically prior to the others, in the sense that justifying it is necessary and sufficient for justifying the others. Then, in the second stage, I will justify this epistemically prior inference. It is here that we will find that all existent definitions of symmetry in the literature leave this inference unjustifiable. To remedy the situation I proceed by reverse-engineering, asking what 'symmetry' could possibly mean if the inference is to be justified. I outline a definition that preserves the original idea behind the concept but which also does the required justificatory work.

The result may be of independent interest to practioners of symmetry reasoning, for the definition I arrive at is quite different from the definitions found in the literature. In particular, while the latter are set out in crisp, mathematical terms concerning the physical structure of the world, I argue that the concept of symmetry can only be adequately understood in mental and counterfactual terms. The formulations found in the literature are, at best, approximations to this concept which abstract from details irrelevant to the author's particular purpose.

I should emphasize at the outset that I will not be attempting to chart out *all* the ways in which the concept of symmetry is used in physics. In particular, I will not discuss the phenomenon of "symmetry breaking" or Noether's theorems relating symmetries and conservation principles (indeed, the former is really quite from the phenomena I will discuss even though it is described with the same name). Instead, my aim is to pick up on one notion of symmetry and examine some core aspects of its use, leaving the question of how those core uses relate to other uses for another time.

I proceed as follows. In Sections 1, 2 and 3, I identify three roles that symmetry plays in physical reasoning, and argue that one of those roles is epistemically prior to the others. In Sections 4 and 5 I ask what the notion of symmetry could possibly mean if it is to legitimately play that epistemically prior role.

1 Introduction to Symmetry

Since this paper is about symmetry, I suppose I should start by telling my reader what I mean by the term. However, answering the question 'What is symmetry?' is the stated aim of Section 5 so I cannot do this until the very end! Indeed, any paper who's aim is to propose an analysis of X inevitably faces the same dilemma: fail to say enough about what X is at the beginning and one will have failed to delineate one's topic; but say too much and one will have completed one's project before one starts.

My route around this dilemma is to start in this section with some paradigm examples of symmetries in physics, and make a few comments about what these examples have in common. Then in the next two sections I will describe the role that the concept plays in physical reasoning. This will give us enough of a functional grip on the concept for us to ask, in the final section, what it could possibly mean if it is to play that role.

So what are some paradigm examples of symmetries in physics? Perhaps the best examples are rigid spatial shifts and uniform velocity boosts. These are symmetries of the laws of classical Newtonian Gravitation Theory (NGT), by which I mean Newton's three laws of motion along with the inverse-square gravitational force law.

But what are rigid spatial shifts and uniform velocity boosts? Unfortunately I cannot say too much without getting into the details of defining 'symmetry'! But here is the rough idea. Consider a classical physical system evolving over time. At each instant of time the system is in a certain state, and we can imagine that the system is composed of particles and that the entire state of the system at each time is determined by the state of each particle—it's mass, charge, position and so forth—and the spatial relations between them at that time. Let us also suppose that the velocity of each particle is a part of its state at each instant.[1] OK, now consider a second system who's state at each time is just like that of the first system but for the fact that every particle is shifted over three feet in some uniform direction. The operation that takes the initial system as input and yields this second system as output is called a rigid spatial shift, and all such spatial shifts (for arbitrary distances and directions of shift) are symmetries of the laws of NGT. Similarly, consider a third system who's state at each time is just like the first but for the fact that the velocity of each particle is five mph greater in a certain specified direction. The operation that takes the initial system as input and yields this third system as output is called a uniform velocity boost, and all such boosts (for arbitrary speeds and directions) are symmetries of the laws of NGT too.

Other paradigm examples include the Lorenz transformations, which are symmetries of the laws of Special Theory of Relativity (STR); and diffeomorphic shifts, which are symmetries of the laws of General Theory of Relativity (GTR). But there is no need to explain what these sorts of transformations are: most of the philosophical points in this paper can be stated in the context of NGT and I will limit myself to that context whenever possible.

From these examples we can already abstract some similarities. First, the concept of symmetry here applies to very different things than our everyday concept of symmetry. In everyday talk we might say that a human body is (roughly) bi-laterally symmetric, or that a plate is symmetric through straight lines that bisect its center. In these cases it is an *object* that is said to be symmetric in a certain way, and in saying this we are saying something about its shape. But the physical concept exemplified above is different: it is used to state that a given set of *physical laws* are symmetric in a certain way. So we might call this a "nomic" or "dynamic" conception of symmetry, rather than the "geometric" conception expressed in everyday talk.

Second, the symmetries themselves, e.g. rigid spatial shifts and uniform velocity boosts, appear to be *transformations* or *operations* on physical systems. For now, I will leave it open exactly what physical systems are: they could be possible worlds, set-theoretic models, or some other entity altogether. But note that a physical system is understood to be not just an instantaneous state but rather an entire history of such instantaneous states (for example, an entire world rather than just a time-slice of a world). One thing I will assume is that it makes sense to talk of a physical system making a given set of laws true. If physical systems are models this amounts to their being models of sentences that express the laws, while if they are possible worlds this amounts to their being worlds in which the laws are true. For

[1]This supposition is entirely dispensable if you think that there is no such thing as instantaneous velocity, but is convenient for illustrative purposes.

our purposes, it does not matter either way.

Third, in all the above cases, the transformations that are said to be symmetries of a given theory all necessarily preserve the truth of the theory. For example, consider rigid spatial shifts, which are symmetries of NGT. These operations have the following property: that when applied to *any* possible physical system that makes NGT true they invariably yield another physical system that also makes NGT true. Exactly the same goes for all the other paradigm cases of symmetry. Indeed, it is uncontroversial to say that a necessary condition for an operation O to count as a symmetry of a given law L is that O necessarily preserves the truth of L.

Not only is this uncontroversial, it is an eminently natural thing to say and nicely connects the nomic conception of symmetry with its geometric counterpart: just as the geometric symmetries of an object are changes to the object's parts that preserve its *shape*, so too the symmetries of a law are changes to physical systems that preserve its *truth*. But I should emphasize that while this condition is necessary, it is not sufficient to capture the way that 'symmetry' is used in physics: the operation that takes each physical system into the system composed of a single particle at rest necessarily preserves the truth of NGT, but is not normally regarded as a symmetry of those laws.

There is something a little perverse about this. Whether geometric or nomic, symmetries appear to be concerned with *preserving* something; and, if so, the symmetries of X should really just be transformations that preserve X. So it is strange to learn that not all transformations that preserve the truth of NGT are counted as symmetries of NGT. As it turns out, when theorists in the literature talk of the 'symmetries of a law L', they actually require the transformation to preserve something else in addition to the truth of L, so the terminology is a little misleading. We will ask what this something else is in Section 5; for now, I just want to introducing our topic intuitively.

Before we move on, I should note that some theorists refer to symmetries as being transformations on *descriptions* of physical systems, rather than on the physical systems themselves. On this approach, a rigid spatial shift would be said to be an operation that takes a coordinate system and delivers another whose origin is shifted over, say, three feet to the right.[2] The analogous necessary condition on such a transformation counting as a symmetry of a given law L is that if the first description describes a system that makes L true then the second one does too. But nothing will turn on the difference between this way of talking and the one I will adopt, on which symmetries are operations on the physical systems themselves.[3]

[2]The distinction between this approach and the one I take in the paper is often referred to as the distinction between interpreting transformations on coordinate systems "actively" or "passively".

[3]I should also note one other variety of approaching symmetry, on which they are

2 Three Roles in Physical Reasoning

So much for our paradigm cases of symmetry. Our next task is to identify the role that the concept plays in physical reasoning. In this section I will distinguish three roles. The roles are closely related and as a result are often conflated, but it will be important to distinguish them. In the next section I will argue that one of them is epistemically prior to the other two, but for now let us just describe the roles without worrying about priority.

To describe them I need a bit of terminology. Consider a physical system and imagine the result of subjecting it to a rigid spatial shift. Notice that the resulting system will be exactly the same as the original system with respect to all facts about *relative positions*: if two materials bodies were two feet apart at a given time t in the original system, they will be two feet apart at t in the resulting system too. And it turns out that if we were to subject the original system to *any* symmetry of NGT, the resulting system would agree on all facts about relative positions. Because of this, I will say that relative position is *preserved by* the symmetries of NGT. Not so with *absolute position*, by which I mean the position of a material body in absolute space: if we subject a physical system to a rigid spatial shift, the resulting system will differ with respect to facts about the absolute position of each material body. Because of this, I will say that absolute position *varies under* the symmetries of NGT. Similarly, *relative velocity*—the velocity of a material body relative to another material body—is preserved by the symmetries of NGT, but *absolute velocity*—the velocity of a material body through absolute space—varies under those symmetries (just think of uniform velocity boosts).

Actually, this terminology needs to be qualified slightly. Some theorists (myself included) think that there is no such thing as absolute position or velocity, and that the laws of NGT are better formulated on a space-time structure that does not contain such features.[4] But they also believe that absolute position varies under the symmetries of NGT (indeed, as we will

operations on *physical states at a given time*, rather than complete systems over time. For these theorists, the necessary condition on such an operation counting as a symmetry of a given law L is that the result of evolving a given state S according to L and then operating is the same as operating on S and then evolving the state according to L. For an example of this approach, see Wigner [14]). However, the difference between defining symmetries on states-at-times and defining them on entire physical systems over time will not matter for our purposes, so I will ignore it in what follows.

[4]I am slurring over a subtlety here. NGT *as written down by Newton himself* make reference to absolute velocity, so what Newton wrote down would not be true if there were no such thing as absolute velocity. But I am using the phrase 'NGT' to refer to laws that can be expressed in different ways depending on what one takes to be the underlying metaphysics of the world they govern. We use the phrase 'the Schrodinger equation' in quantum mechanics the same way, to refer to a law that will ultimately be formulated in very different ways depending on one's view about the ontology of a quantum mechanical world.

see this latter belief can be used to motivate the former belief). So it is more accurate to put things like this: absolute position varies under the symmetries of NGT because *if there were such a thing as absolute position*, some of the symmetries of NGT (e.g. rigid spatial shifts) would transform physical systems into systems that differ with respect to facts about absolute position. This is true even if there is no such a thing as absolute position.

Let me now describe three roles that symmetry plays in physical reasoning.

Role 1: Symmetry as a Guide to the Undetectable

The first role uses the symmetries of a given set of laws to teach us about what sorts of features are *undetectable*—or, at least, what features would be undetectable were those laws to be the true and complete laws governing our world. To see this role in action, here is a quote from Earman:

> Because Newton's laws of motion and gravitation have (Gal) as their dynamical symmetries [i.e. because the symmetries of NGT include uniform velocity boosts], no feature of the lawlike behavior of gravitating bodies can be used to distinguish an absolute fame: in this sense, absolute space is unobservable.[5]

Now, it is is reasonably clear from the context that by 'absolute space' he is referring to our *velocity through* absolute space. Moreover, by saying that absolute space is 'unobservable' it is reasonably clear that he does not just mean that absolute velocity is invisible to the naked eye, where this leaves it open that it is detectable by means of complex measuring devices. Rather, he means that it is impossible to observe or measure or otherwise determine our velocity through absolute space by any process of investigation at all. So it is perhaps better to rephrase him as concluding that absolute velocity is *undetectable* rather than unobservable.

Actually, he does not really conclude with that either, since he is only reasoning in the context of NGT. So we should really see him as concluding that absolute velocity *would* be undetectable *were* NGT true and complete. (However, for ease of prose I'll sometimes leave the conditional nature of this conclusion implicit.) Notice that Earman draws this conclusion *purely* on the basis of the fact that the symmetries of NGT include uniform boosts, which are transformations that vary absolute velocity. So Earman's quote is an instance of the following form of inference: one infers from a premise of the form

(1) Feature X varies under the symmetries of a given set of laws L

to a conclusion of the form

[5] Earman, [4], p. 48.

(2) Feature X would be undetectable if L were the true and complete laws of motion.

Once pointed out, this form of inference can be seen all over the place. Earman's quote above is one example, but for another consider this quote from Huggett:

> The equivalence, with respect to Newton's laws, of frames in constant relative motion is at the root of the undetectability of absolute velocity.[6]

Here, the equivalence he refers to is just the fact that uniform velocity boosts are symmetries of NGT. And, like Earman, he seems to think that we can derive on this basis the conclusion that absolute velocity is undetectable. For another example, here is Feynman:

> ...the laws of Newton are of the same form in a moving system as in a stationary system, and therefore it is impossible to tell, by making mechanical experiments, whether the system is moving or not.[7]

By saying that the laws of NGT are "of the same form in a moving system as in a stationary system", it is clear from the context that he means uniform velocity boots are symmetries of those laws. So here is yet another example in which a theorist infers directly from a premise about what features vary under the symmetries of a set of laws, to the conclusion that that feature is undetectable.

The first role of symmetry I want to highlight, then, is its role in the inference from (1) to (2). But before we move on, I want to point out a very obvious but underappreciated fact about all the quotations above, namely that none of the authors provide any kind of justification for this form of inference! Take a moment to appreciate the leap made when one infers from (1) to (2): from what has been said so far about symmetry, it is totally unobvious why facts about the symmetries of a theory—i.e. facts about what transformations on physical systems necessarily preserve the truth of the theory—should imply *anything* about epistemic facts concerning what it is possible for us to detect.[8] So they must be making an implicit assumption connecting symmetry with detection:

[6] Huggett [7], p. XX

[7] Feynman [5], p. 15-2.

[8] To be fair, Earman does spend a good amount of time defining the concept of symmetry before that quote. However, as we will see in Section 4 the definition that he works with does nothing to justify the inference. So the impression I am giving in the text—of Earman drawing this conclusion purely on the basis of what *I* have so far said about symmetry—is a bit unfair, but by the end of the paper I will have argued that it gives an accurate impression of the state of play.

The Symmetry-Detection Link Necessarily, if the laws of L are true and complete, then any feature that varies under the symmetries of L is undetectable.

If this principle is true, then the inference from (1) to (2) goes through; if it isn't, then it doesn't. Thus, the legitimacy of the form of inference stands or falls with the Symmetry-Detection Link.

It is surprising to discover, then, that there are almost no explicit statements of the Symmetry-Detection Link in the literature![9] This should be extremely puzzling: instances of the form of inference from (1) to (2) are, as I have argued, made throughout the literature with very little justification, but one almost never sees an explicit statement—let alone justification—of the principle underwriting them.

It is worth pausing to appreciate how powerful the Symmetry-Detection Link is. If it is true, then we could take our best-confirmed physical theory, ask what its symmetries are, and then draw conclusions about the limits of what we can detect. And the results could be surprising. Remember, until recently the belief that the earth is at rest in space was central to our world-view. Why? Presumably because to the naive eye that is exactly what it looks like—when I look at the buildings and trees around me they look totally still! But with the Symmetry-Detection Link in hand we could have discredited this world-view by (i) gathering evidence to confirm NGT, (ii) working out that uniform velocity boosts are symmetries of NGT, and finally (iii) inferring that the earth's velocity through space is, surprisingly enough, undetectable. This conclusion would have flabbergasted our ancestors, but if the Symmetry-Detection Link is true it can be justified by simple experimental evidence confirming NGT.

In sum: the first role of symmetry is its role in the inference from (1) to (2). This form of inference is valid if and only if the Symmetry-Detection Link is true, so we can equivalently see the first role of symmetry as being its role in that principle.

Role 2: Symmetry as a Guide to Reality

The first role of symmetry was to inform us of what is detectable. The second role is to inform us of what is *real*. When one reasons in this way, one infers from a premise of the form

(3) Feature X varies under the symmetries of a given set of laws L

to a conclusion of the form

(4) Feature X is not real (or wouldn't be, were L the true and complete laws of motion).

[9] Roberts is the only theorist I know of who states it explicitly; see his [12].

This form of inference can be found throughout the physics and philosophy of physics literature. For example, Earman explicitly uses it when arguing that if our world were Newtonian we should think that there is no such thing as absolute velocity.[10] And North explicitly uses it when asking metaphysical conclusions we can draw by examining Hamiltonian and Lagrangian formulations of classical mechanics.[11]

Those are two examples in which the inference from (3) to (4) is explicit, but it is also implicit in physical reasoning. To see this, consider the Special Theory of Relativity (STR). It is often remarked that according to the laws of STR there are no facts about absolute simultaneity, i.e. facts about whether two events occur at the same time; the idea being that only facts about simulteneity *relative to an inertial frame of reference* are real. But this remark is not strictly speaking true: on the Lorenz formulation, the laws of STR govern the motions of particles through a classical space-time structure in which there are well-defined facts about absolute simultaneity. What *is* true, though, is that absolute simultaneity varies under the symmetries of the laws of STR, and the form of inference from (3) to (4) would have us conclude on this basis that there is no such thing as absolute simulteneity. So the incorrect remark is so often made, I think, because this form of inference is so often implicitly made.

Once again, it is worth noting that while this form of inference is often made, it is rarely defended. Since it is far from trivial, let us state the sort of principle must be implicitly assumed in order for the inference to go through:

The Symmetry-Reality Link Necessarily, if the laws of L are true and complete, then any feature that varies under the symmetries of L is not real.

Actually, we should distinguish two different attitudes we might have to this principle: one in which we take it to state a truth, and another in which we take it to embody a methodological principle. On this latter attitude, adoption of the principle is not a belief that it states a truth, but rather the belief that the fact that a (putative) feature varies under the symmetries of L is a good reason (though perhaps not a decisive reason) to think that the feature is not real—or, at least, that it would be a good reason if we thought that L was true and complete.[12]

So the second role of the concept of symmetry is its role in the inference from (3) to (4). Since this form of inference is valid if and only if the Symmetry-Reality Link is true (or, if you prefer, since the inference is

[10] See Earman [4].

[11] See North [10] and [11].

[12] Earman's principle [SP1] on p. 46 of his [4] is, in effect, a statement of the Symmetry-Reality Link in this methodological form.

reasonable if and only if the Symmetry-Reality Link is a reasonable methodological principle), we can equivalently see this second role of symmetry as being its role in that principle.

Role 3: Symmetry as a Guide to Theory

Until now I have imagined that we start out with well-confirmed beliefs about what the laws of motion are, and then sit down and work out what their symmetries are; and only then, finally, do we use the two links just discussed to infer which features are undetectable or unreal. But we can also reason in reverse, using symmetry as a constraint on an adequate physical theory.

Of course, one way to do this is just to use the Symmetry-Detection Link or the Symmetry-Reality Link contrapositively. For suppose we started off with the conviction that a certain feature—say, shape—was real and detectable. If that conviction were true, it would follow straight from either of (the contraposition of) the two Links discussed above that the symmetries of the true and complete laws of motion will preserve that feature. Thus, from antecedent convictions about what is real or what we can detect, we can infer that it is a condition of adequacy on a physical theory that its symmetries preserve that feature.

But there is another way that symmetry considerations can be used as a guide to theory. To see this, consider Einstein's derivation of his Special Theory of Relativity (STR). He started out with two principles—the special principle of relativity and the principle that the speed of light is independent of the speed of its source—and from these he derived the core features of STR such as length contraction, time dilation and so on. Now, what exactly is the special principle of relativity? In his 1905 paper, Einstein expressed it like this:

> The laws by which the states of physical systems undergo change are not affected, whether these changes of state be referred to the one or the other of two systems of coordinates in uniform translatory motion.[13]

He later gave it a slightly clearer expression:

> If a system of coordinates K is chosen so that, in relation to it, physical laws hold good in their simplest form, the *same* laws also hold good in relation to any other system of co-ordinates K' moving in uniform translation relatively to K. This postulate we call the "special principle of relativity".[14]

[13] Einstein [2], p. 4.
[14] Einstein [3].

In both quotations he talks of transformations on coordinate systems that preserve the truth of the laws. But as I said in Section 1, this is just one way to talk about symmetries. So the more general expression of the principle is that uniform velocity boosts are symmetries of the laws. Stating this explicitly as a constraint, we get:

> **The Special Principle of Relativity (First Version)**: Whatever the true and complete laws of motion are, they will include uniform velocity boosts amongst their symmetries.

As I said, along with the principle that the speed of light is independent of the speed of its source, Einstein used this principle of relativity to derive the invariance of the laws under Lorenz transformations, and the phenomena of time dilation, length contraction, and the relativity of simulteneity. Indeed, Einstein used symmetry considerations in a similar way when formulating his GTR. In that case, he appealed to a *General* Principle of Relativity to constrain an adequate physics, and (along with other constraints) derived the core features of GTR. The details of this second case will not matter to us in what follows, all that matters is in both cases he used symmetry as a guide to theory. So this is the third role of symmetry in physical reasoning: it can be used to formulate constraints on an adequate physics which, along with other constraints, guide the discovery of physical theory.

Given my nomenclature it will not surprise the reader that the special principle of relativity has also been expressed differently, and it will be useful to have the second version on the table. This second version usually finds its expression in more popular expositions of STR. Here is one example from an undergraduate text-book:

> It is impossible to devise an experiment that determines whether you are at rest or moving uniformly.[15]

The text-book calls this the "Principle of Relativity", and it then proceeds to use it (along with the principle concerning light) to derive the core features of STR. Similarly, Feynman says this when introducing STR:

> . . . if a space ship is drifting along at a uniform speed, all experiments performed in the space ship and all the phenomena in the space ship will appear the same as if the ship were not moving. . . That is the meaning of the principle of relativity.[16]

Notice that these statements of the principle of relativity do not saying anything explicit about laws or symmetry. Instead, they express the idea that we cannot tell different states of uniform motion apart from one another. In our terminology of undetectability, the idea becomes:

---

[15] Tipler [13], p. R-1.

[16] Feynman [5].

The Special Principle of Relativity (Second Version): Absolute velocity is undetectable.

So here we have two "principles of relativity", each of which has been taken to imply (along with the principle concerning light) the core features of STR. I will discuss the relation between the two principles in the next section. In particular, I will argue that the First Version is actually indispensable from any derivation of STR, i.e. that the second version does not suffice on its own. But put this aside for now, for even if one *can* derive STR from the Second Version, it remains the case that one *can* also derive STR from the First Version just as Einstein did. This is all I need to make my current claim, which is that symmetry can be (and indeed has been) used as a guide to theory.

3 The Primacy of the Symmetry-Detection Link

So those are three roles that the concept of symmetry plays in physical reasoning. They may not be the only roles, but they are enough to think about for now. The question I want to turn to is: are they justified?

As I mentioned in the introduction, it is a striking fact about the physics and philosophy of physics literature that this question receives very little attention! And when it does, the justifications on offer are woefully inadequate (we will see some examples later in this section). My main aim in this paper is to try to fill this lacuna and provide a "rational reconstruction" of these three roles.

In this section I will argue that the second two roles—our use of symmetry as a guide to reality, and our use of symmetry as a guide to theory—can be justified only by appeal to the Symmetry-Detection Link. In this sense, the first role—our use of symmetry as a guide to detection—is epistemically prior to the other two. Of course, this is not to say that the Symmetry-Detection Link is *metaphysically* prior to the other two, just that a rational reconstruction of the other two roles must start with a justification of the Symmetry-Detection Link. Then in the following two sections I will try to justify the Symmetry-Detection Link itself.

Occamist Reasoning

Let me start with the second role that symmetry plays in physical reasoning. That second role, remember, is underwritten by:

The Symmetry-Reality Link Any (putative) feature that varies under the symmetries of the true and complete physical laws of motion, whatever those laws turn out to be, is not real.

I mentioned earlier that the Symmetry-Reality Link could be understood as a truth or as a methodological principle. Let us consider its status as a methodological principle, though analogous points can be made if it is understood as expressing a truth. I will argue that the best justification of this principle rests on the Symmetry-Detection Link.

Let us start by asking how the principle is justified in the literature. I said before that it is largely impicit and therefore has received little justification, but on the few occasions it has been discussed explicitly the justification appears to be Occamist. For example, here is Earman discussing a version of it restricted to geometric features of space-time:

> The motivation [for this principle] derives from combining a particular conception of the main function of laws of motion with an argument that makes use of Occams razor. Laws of motion, at least in so far as they relate to particles, serve to pick out a class of allowable or dynamically possible trajectories. If [there are geometric features that vary under the symmetries of the laws], the same set of trajectories can be picked out by the laws working in the setting of a weaker space-time structure. The theory that [accepts these features] is thus using more space-time structure than is needed to support the laws. . . [17]

For Earman, then, the Symmetry-Reality Link is justified on Occamist grounds. The idea seems to be that if a feature such as absolute velocity varies under the symmetries of our laws, the particular values it takes are of no relevance to whether or not the laws hold: after all, there are operations on its values, like uniform velocity boosts, such that given any physical system that makes the laws true, the result of changing the values in accordance with the operation invariably yields another system that conforms to the laws. So the feature is said to be "superfluous" to the truth of the laws, and Occam's razor has us dispense with it.

But on the face of it, this is an insufficient justification: Occam's razor does not tell us to dispense with structure that is redundant to the *truth of the laws*; it tells us to dispense with structure that is redundant to explanations of what we *see and detect*. So, for example, if we could really see or otherwise detect our velocity through absolute space, surely we would have a great reason to think that absolute velocity was real! The fact that absolute velocity varies under the symmetries of the laws of motion would then be of little relevance when working out what features are real.

But here is where the Symmetry-Detection Link comes to the rescue: for if it is true, the fact that absolute velocity varies under the symmetries of NGT implies that it would be impossible to detect absolute velocity if NGT were true and complete. Given this implication, it would appear that

[17] Earman [4], p. 46.

we can apply Occam's razor after all: for if the particular values a feature takes is superfluous to the truth of one's best physics, *and if in addition the feature is undetectable*, it would seem reasonable to think on Occamist grounds that the feature is not real—at least, all else being equal.[18] To read the above quote from Earman charitably, then, I suggest that we interpret him as implicitly assuming the Symmetry-Detection Link.

I said two paragraphs back that the Symmetry-Reality Link cannot be justified merely by appeal to Occam's razor, but one might object to this on the grounds that, when taken as a methodological principle, the Link only ever purported to give us *some* reason to think that features that vary under the symmetries of the laws are unreal, and surely Earman's Occamist consideration is enough to give you *that*. Sure, the objector will say, this reason is obviously defeated if one can detect the feature after all, but that is consistent with the Occamist consideration giving us *some* reason to dispense with it, just as Earman says.

In response, I agree that one *could* interpret the Symmetry-Reality Link as giving us a reason to dispense with the feature that is hostage to further facts about what we can detect. But this weak version of the Link is not the version that has been used (explicitly or implicitly) in the literature. For with only this weak version in hand, one could never legitimately use the fact that a feature varies under the symmetries of the laws to argue that it is unreal, without *also* engaging in an investigation into whether the feature is detectable. But this is exactly how theorists have argued! For example, when Earman dispenses with absolute velocity, he does not discuss whether absolute velocity is detectable but instead takes the fact that it varies under the symmetries of the laws to provide sufficient justification to dispense with it.[19] And North does the same when dispensing with various metric features.[20] Given the way these theorists actually reason, then, they are clearly taking the fact that a feature varies under the symmetries of the laws to give us a reason to dispense with the feature that is *not* hostage to further facts about what we can detect. My objection is to taking the Link to have *this* strength and then justifying it merely by appeal to Occam's razor: without the Symmetry-Detection Link, Occamist considerations alone simply do not yield a Link of that strength.

So this is why I think that the Symmetry-Detection Link is needed to justify the Symmetry-Reality Link. *Contra* what Earman explicitly states,

[18]And of course all else is *not* always equal, which is presumably why we sometimes (rationally) believe in features that vary under the symmetries of the physical laws. Newton, for example, believed in absolute velocity because he did not know how to formulate an empirically adequate theory without it, and I would argue that contemporary substantivalists believe in a manifold of spatio-temporal locations for the same reason. Nonetheless, they can all agree that there is at least *some* reason to dispense with features that vary under the symmetries of the laws, even if that reason is not decisive.

[19]See Earman [4].

[20]See North [10] and [11].

the mere fact that a feature like velocity varies under the symmetries of the laws is, in and of itself, of little relevance when figuring out what is real. The beauty of the Symmetry-Detection Link is that it allows us to turn this fact into a good reason to dispense with the feature after all.

Relativity Principles

I will now turn to the third role that symmetry plays in physical reasoning and argue, again, that it can be justified only by appeal to the Symmetry-Detection Link. That third role, remember, was the use of symmetry as a guide to physical theory. I described the case in which Einstein took the following as a constraint on physical theory

> **The Special Principle of Relativity (First Version)**: Whatever the true and complete laws of motion are, they will include uniform velocity boosts amongst their symmetries.

and, along with the principle concerning light, derived the core features of STR. I will argue that his use of this constraint on physical theory is justified only by appeal to the Symmetry-Detection Link.

But I will argue for more than this. I said earlier that some text-books claim to derive the same features of STR by substituting this First Version with the following:

> **The Special Principle of Relativity (Second Version)**: Absolute velocity is undetectable.

Since this principle does not contain explicit reference to symmetry, one might think that STR can be derived without engaging in symmetry reasoning at all. But I will argue that this is a mistake: even though the Second Version does not explicit make reference to symmetry, its use in guiding physical theory is justified only by appeal to the Symmetry-Detection Link too. If all this is right, then the Symmetry-Detection Link is a core part of *both* these ways of deriving STR.

Start with the First Version. It says that it is a condition of adequacy on any physical theory that its laws include uniform velocity boosts amongst their symmetries. But what justifies us putting this constraint on physical theorizing. What, exactly, would be wrong with believing that our world is governed by laws that violate the principle?

The answer, I think, lies in the Symmetry-Detection Link. To see why, it will help to consider a passage from Einstein in which he is justifying a related principle, his General Principle of Relativity. This principle says that the true and complete laws of motion must include a different set of transformations amongst their symmetries. Never mind what those transformations are, all we need to know is that Einstein called this the requirement of "general covariance". Now, what would justify such a constraint on physical theorizing? Here is Einstein:

> All our space-time verifications invariably amount to a determination of space-time coincidences. If, for example, events consisted merely in the motion of material points, then ultimately nothing would be observable but the meeting of two or more of these points... coincidences between the hands of a clock and the points on a clock dial, the observed point-events happening at the same place at the same time.
>
> ... As all our physical experience can be ultimately reduced to such coincidences, there is no immediate reason for preferring certain systems of co-ordinates to others, that is to say, we arrive at the requirement of general co-variance.[21]

Now, the precise interpretation of this passage is a matter of considerable controversy in the history of physics.[22] But one thing is reasonably clear: he appears to be justifying the requirement of general covariance by appealing to what we can *observe*. Moreover, the facts that he says we can observe—facts about space-time coincidences—are all facts that are preserved by the symmetries of generally covariant theories. So, the analogue of this idea in our more familiar case of STR is this: what justifies The Special Principle of Relativity (First Version) is an antecedent belief that we cannot detect absolute velocity. On this suggestion, then, the First Version is actually justified by appeal to the Second Version!

So far so good. But why exactly would an antecedent belief that absolute velocity is undetectable justify us in concluding that the true and complete laws of motion include uniform velocity boosts, as the First Version states? Well, if the Symmetry-Detection Link is assumed, then the hypothesis that the symmetries of the laws of motion include uniform velocity boosts would explain the fact that absolute velocity is undetectable. So I suggest that the inference from the Second Version to the First Version is best thought of as an Inference to the Best Explanation—an explanation that assumes the Symmetry-Detection Link. If that is right, then our justification for using the First Version as a constraint on physical theorizing appeals to the Symmetry-Detection Link.

In defense of this claim, let me object to two other ways in which the Second Version might be taken to justify the First Version. First, one might suggest that it is justified by the converse of the Symmetry-Detection Link, i.e. the claim that all undetectable features vary under the symmetries of the true and complete laws of motion. Along with this claim, the inference from the Second Version to the First Version is logically valid. But I reject this justification because I see little reason to believe the converse of the Symmetry-Detection Link. After all, a given feature might be undetectable for many reasons. For one, the feature may only be instantiated

[21]Einstein [3] p. 177.

[22]See Howard [6] for a good overview.

in environments that are utterly hostile to the existence of any cognitive life—say, near the singularity of a black hole—and may not have any effects on other regions. For another, the feature may be instantiated around us but may require more energy to detect than we can possibly harness. Of course, if the Symmetry-Detection Link is true then *another* reason that a feature might be undetectable is that it varies under the symmetries of the true and complete laws of motion. But since there are many reasons why a feature might be undetectable, I see little reason to accept the converse of the Symmetry-Detection Link.

Second, one might think that the inference from the Second Version to the First Version proceeds as follows: from the Second Version (which says that absolute velocity is undetectable) we conclude by Occam's razor that absolute velocity is not real, from which it follows that any law must be symmetric under uniform velocity boosts (on pain of the law's truth depending on a feature that doesn't exist), just as the First Version says. But the problem with this justification is that it is not sufficiently general. Lorenz, for example, formulated a version of STR in which the laws governed particles moving with perfectly real and determinate velocities through absolute space. Since he thought that absolute velocity was real, the current suggestion would preclude him from having used symmetry considerations as a guide to physical theory. Therefore, the current suggestion is unreasonably restrictive.

For these reasons I prefer my own reconstruction of the inference from the Second Version to the First Version. On this account, it is not a valid inference but an inference to the best explanation which assumes the Symmetry-Detection Link. The idea is that absolute velocity is undetectable, and this is best explained (if the Link is true) by the thesis that uniform velocity boosts are symmetries of whatever the true and complete laws of motion are.

That completes my argument for the claim that our use of the First Version to constrain physical theorizing can be justified only by appeal to the Symmetry-Detection Link. As I said, I also claim that our use of the Second Version to constrain physical theorizing is *also* justifiable only by appeal to the Symmetry-Detection Link. In fact, it should now be easy to see that the Second Version is not enough to constrain physical theorizing on its own. For the Second Version just says that absolute velocity is undetectable, and we have already seen that there are many reasons why a feature might be undetectable: it might be manifested only in hostile environments, or it might require too much energy to detect it. In either of these two cases, it is perfectly consistent with the undetectability of the feature that it is *preserved* by the symmetries of the true and complete laws of motion. So, strictly speaking, it is perfectly consistent with the Second Version that absolute velocity is preserved by the symmetries of the true and complete laws of motion. And if absolute velocity is preserved by those symmetries, then the Lorenz transformations are not symmetries of the true and complete laws of motion (since absolute velocity varies under their symmetries). But one of the things that the principle of relativity and the principle about the speed of light are supposed to imply is that the true and complete laws of motion must be symmetric under Lorenz transformations! Therefore, if we take the principle of relativity to be *merely* that absolute velocity is undetectable, as the Second Version does, this is not sufficient to constrain physical theory in the way that the principle is taken to.[23]

Why, then, do so many theorists claim to derive STR from Second Version? I think it is because when these theorists write down the Second Version, they have at the back of their mind not just the idea that absolute velocity is undetectable but that it is undetectable for a particular reason: namely that it varies under the symmetries of the true and complete laws of motion. If that is right, then their derivation of STR just collapses into the first. Once again, the structure is this: we start with the antecedent belief that absolute velocity is undetectable; we infer to the best explanation that this must be because the true and complete laws include uniform velocity boosts amongst their symmetries; and we use *this* constraint to derive the core features of STR.

In sum: our use of the First Version can only be justified by appeal to the Symmetry-Detection Link, and our use of the Second Version is unjustifiable on its own and, once justified, collapses into a use of the First Version. Either way, the Symmetry-Detection Link plays a central role in justifying our use of symmetry as a guide to theory.

4 Justifying the Symmetry-Detection Link

So far I have argued for the (perhaps unsurprising) claim that symmetry plays a central role in physical reasoning, and the (perhaps more surprising) claim that those roles are underwritten by the Symmetry-Detection Link. But is the Symmetry-Detection Link true? As it turns out, the literature provides almost no discussion of this question! So I take it to be a pressing concern in the philosophy of physics to provide such a justification. In this section and the next I will attempt just this.

Now, while the literature contains very little discussion of the Symmetry-Detection Link in general, one finds some discussion of one particular in-

[23]Of course, I have not proven that the principle concerning light does not close the gap here, but it is hard to see how it possibly could if we are seriously leaving open the possibility that the only reason why absolute velocity is undetectable is, say, that we cannot harness enough energy to uncover its values. Another way to argue the point is this: the principle of relativity is often said to imply by itself that the true and complete laws are either Lorenz or Galilean invariant (with the principle concerning light ruling out the latter disjunct), but clearly the mere fact that absolute velocity is undetectable does not imply this disjunction.

stance: that if the laws of NGT are true and complete then absolute velocity is undetectable. So in this section I will outline the best argument for this instance I can glean from the literature and then ask whether the argument generalizes to arbitrary laws and features. It will be clear that the answer depends on what we mean by 'symmetry', so in the next section I will try to find a definition of the term on which the answer is 'Yes'.

The Galilean Two-Step

It is often said that if the laws of NGT are true and complete, then absolute velocity is undetectable. This seems absolutely right; but how, precisely, does the undetectability follow from those laws being true and complete? More precisely, how does it follow from facts about the *symmetries* of NGT, as the Symmetry-Detection Link states?

Unfortunately, one finds only passing remarks in the literature. For example, Maudlin writes that 'according to Newtonian physics itself, no mechanical experiment could determine the absolute velocity of a body. Laboratory experiments done in a room at absolute rest would have identical outcomes to those done in a room moving with constant velocity'.[24] Elsewhere he writes that 'since only acceleration has any dynamical consequences [according to Newtonian mechanics], uniform absolute motion would produce no observable change'.[25] Helpful though these remarks are, one isn't told how the argument in favor of these claims are supposed to go, whether they appeal to facts about the symmetries of NGT, and (most importantly) whether the argument easily generalizes to arbitrary laws and features, as it would need to if it is to justify the Symmetry-Detection Link.

A recent paper by John Roberts fills in some of these gaps. He outlines an argument for this instance of the Symmetry-Detection Link that appears to appeal to nothing other than facts about the symmetries of NGT, and which therefore holds the promise of generalizing. I will reconstruct his reasoning in some detail here, since it will help to bring to light what must be the case if this form of argument is to generalize.[26]

To start, imagine trying to build a device that detects whether it is moving or at rest. What properties must this device have? At a minimum, it would need to be sensitive to whether it is moving or not. But it is worth being quite specific about what this means. Presumably, it would need to have a "ready state" in which it is set up to measure its velocity through space. And it would need to be built in such a way that if it were in its

[24] Maudlin [8], p. 40.

[25] Maudlin [9], p. 189.

[26] Roberts sets the argument out in his [12]. I should say that while the following argument is largely borrowed from Roberts, my presentation here departs in some ways from his specific way of running the idea. I have also heard this kind of argument given in seminars by Tim Maudlin and David Albert.

ready state at some initial time t_0 and was then switched on *while at rest*, it would whir away and then register this fact by, say, displaying 'At Rest' on a computer screen at some later time t_1. Correspondingly, if it were in its ready state at t_0 and then switched on *while moving*, it would then register this differently, say by displaying 'Moving' on the screen at t_1 instead.

But it turns out that if NGT were the true and complete laws governing our world, it would be impossible to build a device with these properties. Moreover, the argument for this claim makes explicit appeal to the fact that absolute velocity varies under the symmetries of NGT. So if the argument is any good, it might suggest an argument for the claim that the Symmetry-Detection Link is true in general.

The argument goes like this. Let us suppose that I succeed in building a device with one of the properties described above: if it were switched on at t_0 while at rest, it would register 'At Rest' on its screen at t_1. Now, my turning the device on and it whirring away and displaying 'At Rest' on its screen constitutes a physical system, so we can consider the result of applying a uniform velocity boost to it. What is the resulting system like? We know three things about it. First, it is one in which at t_0 the device is turned on *while moving*. Second, since relative positions are preserved by uniform velocity boosts, it is also a system in which at t_1 the word 'At Rest' appears on its screen, just like in the original system. Third—and here is the crucial point—since uniform velocity boosts are symmetries of NGT, it follows that the resulting system conforms to NGT just like the original system (remember, a necessary condition on an operation being a *symmetry* of a set of laws is that it necessarily preserves the truth of those laws). It follows from this third point that the resulting system represents *how the device would evolve were it to be turned on while in a state of motion*. Therefore, if NGT were true and complete, then given *any device whatsoever* that is built to display 'At Rest' when turned on at rest, it follows straight from the fact that uniform boosts are symmetries of NGT that it would display 'At Rest' when turned on while moving too. And so it follows that if NGT were true and complete it would be impossible to build a device with the property specified two paragraphs back. QED.

Notice how powerful this form of argument is: analogous reasoning establishes that if our world were Newtonian, no device could display our absolute velocity in pointer positions, or patterns of ink on computer printouts, or sounds emanating from speakers, or.... And it is hard to see how else a device could display our velocity, if not in these sorts of media. Indeed, something stronger follows: since uniform velocity boosts preserve all facts about relative positions, the same form of argument establishes that no device could display our absolute velocity in any medium who's states supervene on the relative positions between particles, for (using the above reasoning) any such medium will necessarily result in exactly the same display whether the device is turned on at rest or in a state of uniform motion.

So we have at least the beginnings of an argument that proceeds from the fact that uniform velocity boosts are symmetries of NGT, to the claim that absolute velocity undetectable in Newtonian worlds—precisely the sort of argument we were looking for.

Now, we were interested in this sort of argument because we were interested in whether it will generalize to the case of arbitrary laws and features. To help us assess this question, let us represent the line of argument as proceeding in two steps. We took a set of laws L (in this case, NGT), and a feature F that varies under the symmetries of L (in this case absolute velocity), and we argued as followed:

Step One On the assumption that the laws of L are true and complete, we used facts about the symmetries of L to work out how physical devices would behave under counterfactual initial conditions that differed only in the value of F.

Step Two On the basis of how physical devices would behave under those counterfactual initial conditions, we argued that it would be impossible for us to detect F.

I will call this form of argument the "Galilean Two-Step".

So, does the argument go through in the general case? It certainly goes through in *some* other cases. For example, consider the case of absolute position. In Step One, an analogous argument to that rehearsed above would convince us that since rigid spatial shifts are symmetries of NGT, it would be impossible (if NGT were true and complete) to build a device that displays different words on a computer system depending on whether the device was switched on in one region of space or another. And in Step Two we would argue that because of this constraint on how physical devices behave, it would be impossible to detect our position in absolute space.[27]

Generalizing

However, generalizing to one other case does not amount to generalizing to them all. One worry is whether the argument generalizes to the case of indeterministic laws, since in the case of NGT Step One appealed to the fact that NGT is deterministic. I believe it can, but the issues here are more technical than philosophical so I propose to bracket this worry for now.[28]

[27] One might worry that Maudlin provided a counterexample to the Symmetry-Detection Link in his [9], in which he argued that our location in absolute space, which varies under the symmetries of NGT, would in fact be *detectable* in an NGT world. Maudlin's arguments are indeed ingenious, but I believe that what they actually teach us is that the interesting epistemic position that Maudlin shows we can get in with regards our location in absolute space does not in fact consist in *detecting* our position. But a discussion of this would take us too far from the main thread.

[28] Briefly, given a feature F that varies under the symmetries of an indeterministic law L, what one must show in Step One is that given any physical device built to measure F,

Instead, I want to focus on the problem of generalizing Step Two. Indeed, in the case of NGT and absolute velocity discussed above I said very little about this step and spent most of my time motivating Step One. Admittedly, Step Two is enormously plausible in the case discussed, but since we are interested in generalizing the argument we need to take care to see exactly how it goes.

To see some potential problems with the step, imagine an interlocuter called Al: "I grant your reasoning in Step One, which concluded that if NGT were true and complete, no device could be built to detect absolute velocity and record it in any medium which supervenes on the relative positions between things. But something has gone wrong in Step Two, for it does not follow that absolute velocity is undetectable. To see this, just consider the device you described that is build to print the words 'At rest' on a sheet of paper if the device is turned on at rest. To be sure, this device would display the same pattern of ink on the paper if it were turned on while in a state of uniform motion, but it does not follow that the outcome state of the device is the same in both cases. Indeed, they are obviously different: in one case those words are *at rest* and in the other case they are *moving*! So, as long as we individuate the device's outcome states to include not just the pattern of ink on the paper but also the ink's state of motion, then this device does a perfectly good job at detecting absolute velocity."

A similar point can be seen by imagining a second interlocuter called Bob: "Forget about Al's nonsense: I will grant you that the outcome states of the device are to be individuated by the pattern of ink on paper, so that the device gives the same outcome whether it is turned on at rest or moving. However, recall that every device has a "ready state", a state in which it is calibrated and ready to perform a measurement. The point of specifying such a ready state is that how the device behaves when it is *not* properly calibrated should not factor into whether it successfully detects the value of a given feature. But then it is open for the designer of this device to say that its ready state requires that it be at rest! If so, then the device will always correctly record the value of our velocity through absolute space *when turned on in its ready state*, and absolute velocity is detectable after all."

No doubt both Al and Bob are making some mistake. But their responses show that even in the relatively straightforward case of NGT and absolute velocity, the move from Step One to the conclusion that absolute velocity is undetectable is far from trivial. Somehow, the reasoning in Step Two has to take facts about how physical devices would behave under various initial conditions in an NGT world, and show us how it follows that absolute

the probability distribution over outcomes of the measurement are the same under all the counterfactual initial conditions that differ only in the value of F. I am slurring over how one is to define the notion of symmetry in the context of probabilistic laws, but as I said I will be bracketing this issue here.

velocity is undetectable.

Roberts himself is aware of this issue and proposes a very natural solution. His idea is that absolute velocity is in some sense not "directly perceptible". So, if we require that the ready states and outcome states of measuring devices are "directly perceptible", we have an immediate response to Al and Bob. If we further assume that everything that is "directly perceptible" supervenes on the relative positions of material bodies, we can easily establish Step Two: we can explain why it follows from facts about how physical devices would behave in an NGT world that absolute velocity is undetectable.

I will argue later that this solution should be rejected. But put that aside for now. What I want to emphasize here is that when working out how best to formulate the Galilean Two-Step we need to take care that our reasoning will generalize to the case of arbitrary laws and features, as is needed if it is to establish the Symmetry-Detection Link. To see the potential problem, let us grant for a moment that Roberts' solution works in the case of NGT and absolute velocity. What makes it work in this case is that the feature that varies under a uniform velocity boost, i.e. absolute velocity, is not directly perceptible. So, in order for this solution to generalize, it must be the case that given *any* set of laws whatsoever, the features that vary under its symmetries are not directly perceptible.

Is this true? Unfortunately, we are not in a position to consider the question, since we have not yet been given a clear definition of 'symmetry' to work with. However, if we were attracted to Roberts' response to Al and Bob in the case of NGT and absolute velocity, we would hope that the correct definition does indeed make it the case that, in general, features that vary under the symmetries of any set of laws are not directly perceptible; and that therefore the defense of Step Two offered above goes through in the general case. Of course, one might not have been attracted to Roberts' response, but the main point stands: if the Symmetry-Detection Link is to be justified in full generality by the kind of reasoning set out in Steps One and Two, then *however* you want to respond to Al and Bob and run Step Two in the case of NGT and absolute velocity, it had better be the case that what you mean by 'symmetry' allows you to run that line of reasoning in the general case.

Thus, we now have two constraints that a definition of symmetry must satisfy, if it is to legitimate the central role of symmetry in physical reasoning: it must (of course) make the Symmetry-Detection Link true, but it must also give us some indication as to how it can be justified. In particular, it must give us some indication as to how the Galilean Two-Step, and Step Two in particular, can be run in full generality.

5 The Meaning of 'Symmetry'

Surprising to learn, then, that all the definitions of 'symmetry' in the literature systematically fail to satisfy these desiderata! Either the definition renders the Symmetry-Detection Link false even in its paradigm cases such as that of NGT and absolute velocity, or else it leaves one facing insurmountable problems when one tries to justify the Link with the reasoning set out in Steps One and Two above. Given the central role that the Symmetry-Detection Link plays in physical reasoning, the task of remedying this problem should be seen as a pressing concern for the philosophy of physics.

In this final section I will survey the existent definitions and argue that they all fail to satisfy our desiderata. Then I will outline an approach to defining 'symmetry' that, I believe, does much better. This will complete my rational reconstruction of the role of symmetry in physical reasoning.

Formal Definitions

I said in Section 1 that symmetries were transformations on physical systems, but I left it open whether a physical system should be thought of as a model or a possible world, or even some other kind of entity. For ease of exposition I will now take them to be possible worlds and therefore think of symmetries as functions on the space of possible worlds. Admittedly, this is not an entirely happy supposition: for example, a uniform velocity boost now becomes a function that takes a world and delivers another world that differs only with respect to facts about absolute velocity, but since many theorists think that there is no such thing as absolute velocity they may deny that there is such a function! So we need to play a little fast and loose here. If you like, think of it like this: if there *were* such a thing as absolute velocity, there would be worlds that differ only in facts about absolute velocity and a uniform boost would be a function mapping worlds to their boosted counterparts. So I ask for a little charity from my reader, but this charity will go a long way to making the exposition smooth.

I also said in Section 1 that a necessary condition for a function F to be a symmetry of a set of laws L is that F necessarily preserve the truth of L. If F is defined on the set of possible worlds, this amounts to the condition that F maps every world in which L is true to a world in which L is true.

That is a necessary condition, but is it also sufficient? We saw in Section 1 that it is not, on pain of falsifying the Symmetry-Detection Link. For consider the constant function F that maps every world to one containing a single particle at rest for all time. This would count as a symmetry of NGT on this definition, but every interesting physical quantity varies under the operation of this function, including spatial distance. So the Symmetry-Detection Link would imply that those quantities are all undetectable in an

NGT world, which is patently false.

Ismael and van Fraassen add the requirement that F is one-one and onto. But this does not help, for there are plenty of one-one, onto functions on the set of possible worlds that necessarily preserve the truth of NGT but which vary quantities that would be detectable in an NGT world such as spatial distance. On their definition these functions would count as symmetries of NGT, so the Symmetry-Detection Link would patently be false.[29]

Ontic Definitions

This suggests that defining the concept of symmetry out of purely set-theoretic apparatus will not do the trick. Instead, let us return to my comment in Section 1 when I said that the symmetries of a law L will preserve not only the truth of L *but also something else* (and are therefore misleadingly referred to as the 'symmetries of L'). What, then, might that something else be?

One might suggest simply defining the symmetries of a given set of laws L to be those functions on possible worlds that necessarily preserve the truth of those laws *and which preserve all detectable features.* This would of course make the Symmetry-Detection Link true; but it would make it a trivial, analytic truth and would therefore not make any sense of the kind of reasoning described in the Galilean Two-Step. It certainly seems as if the Galilean Two-Step is getting at some deep reason as to *why* the Symmetry-Detection Link holds, and we would like our definition of symmetry to tell us exactly what that is.

To find what it is that symmetries must preserve (in addition to the truth of the laws), let us examine the paradigm cases of symmetry for clues. Looking at those cases, we find that while some symmetries vary absolute positions (e.g. the rigid translations of NGT), and while others vary spatial distances (e.g. the Lorenz transformations of STR), all of them preserve a deeper level of physical structure. For example, they all preserve topological structure: if the surface of a give physical body had the topology of a Euclidean plane in a given world W, it would also have that topology in F(W) for any of our paradigm symmetry F. One might naturally think, then, that we should require symmetries to preserve topological structure of space-time. I just mentioned topological structure, but a similar line of thought might

[29] I should say that I do not really have anything but a verbal dispute with Ismael and van Fraassen. They recognize (and indeed emphasize) that their definition will not by itself play the roles I described in Section 2, so they instead introduce another concept to play that role (more on which below). So we all agree that one-one, onto functions that necessarily preserve the truth of the laws are not fit to play a certain role; our difference (so far) is just that while I reserve the word 'symmetry' for whatever plays that role, they reserve it for one-one, onto functions that necessarily preserve the truth of the laws. As it happens, I do not think the concept they introduce to play the special role is fit for the task, but I will discuss that substantive disagreement later on.

be had with other (or more specific) kinds of space-time structure such as continuous structure, differentiable structure, diffeomorphic structure, and so on. That is, we might notice that our paradigm cases of symmetry appear to preserve that structure, and then require that symmetries in general must preserve it too. The result of any such abstraction will be what I call an "ontic" definition of symmetry: a definition that picks a certain bit of physical structure and decrees that symmetries must preserve *that*.[30]

But there is a very different approach to symmetry. This approach notices that in all the paradigm symmetries such as rigid spatial shifts and uniform velocity boosts, the experiences of agents in the transformed world would be qualitatively indistinguishable from the experiences of agents in the original world. Abstracting, we might suggest that, in general, symmetries must preserve (not any particular bit of physical structure but) the kinds of experiential states enjoyed by subjects. I call any definition of symmetry along these lines a "mentalistic" definition.

I believe that only a mentalistic definition of symmetry can make sense of the role of symmetry in physical reasoning. But since *every* definition of symmetry I have seen in the literature is ontic, I should first say why I reject ontic definitions before saying why mentalistic definitions are much more attractive. So, in order to argue against ontic definitions, I will consider two specific ontic definitions and argue that they both fail. These considerations will then point us towards a general argument against all ontic definitions.

To start, let us consider a suggestion of Earman's. He defines the symmetries of a set of laws L to be all and only those functions F that necessarily preserve the truth of L and which also preserve the diffeomorphic structure of space-time.[31] This is clearly an ontic definition, but is it adequate? One might try to argue that would not make the Symmetry-Detection Link true, since there might be laws whose symmetries (on this definition) would vary detectable structure, but I do not want to pursue that line of objection here so let us suppose for the sake of argument that it does indeed make the Link true. Instead, what I want to emphasize is that Earman builds it into the definition of symmetry that diffeomorphic structure is preserved by all the symmetries of any law of motion. So, if we accepted his definition it would follow from the very definition of symmetry that we could never look at the symmetries of our best physics, notice that diffeomorphic structure varies under them, and then use the Symmetry-Detection Link to conclude that that structure is undetectable or use the Symmetry-Reality Link to conclude that that structure is unreal. As I will put it, Earman's definition renders diffeomorphic structure entirely immune to the sorts of "symmetry consid-

[30] Strictly speaking, ontic definitions do not need to be *motivated* by the "abstraction" method just described. So long as the definition picks out a bit of physical structure and decrees that the symmetries must preserve it, I count it as ontic no matter how it is motivated.

[31] See Earman [4], p. 44.

erations" outlined in Section 2. This should strike us as very odd: while we have reasoned in precisely these ways about absolute position, absolute velocity and absolute simultaneity, it is built into Earman's definition that we could never reason this way about diffeomorphic structure no matter what our laws turned out to be. But what is so special about diffeomorphic structure, that would make it impossible for us to ever legitimately reason in this way? Without an answer to this question, the definition appears gerrymandered and *ad hoc*.

This suggests the following general objection against ontic definitions: all ontic definitions are committed to the existence of some sort of structure that is in principle immune to the symmetry considerations, but it is highly doubtful whether any such structure exists.

To appreciate why it is doubtful, let us consider perhaps the most promising suggestion as to what sort of structure is immune to the symmetry considerations. The suggestion I have in mind is that, in order to determine which bit of physical structure must be preserved by symmetries, we must first determine what bits of physical structure are "directly perceivable" and then determine what those physical structures supervene on. Once we settle on that supervenience basis S, we can say that the a function F is a symmetry of L just in case F necessarily preserves the truth of L and also preserves the physical structures S. The idea is that there will then be nothing gerrymandered or *ad hoc* about the suggestion at all: the reason why the structure in S is immune to the symmetry considerations outlined in Section 2—the reason why we can *in principle* never use facts about the symmetries of the laws to conclude that S is undetectable or unreal—is that S subvenes all that is directly perceivable.[32]

This approach to ontic definitions is based on the idea that all detectable properties are either directly perceivable or else detectable based on their effects on directly perceivable properties. This idea has been tremendously influential in philosophy, so it is no surprise that many theorists of symmetry have thought about symmetry in this way. Indeed, this is the approach taken by Ismael and van Fraassen.[33]

Attractive though this approach is, I believe it is hopeless. The problem arises when we ask what the "directly perceivable" properties are. Of course, many have argued that there is no principled basis for distinguishing the "directly perceivable" features from other features, but I want to put that worry aside and raise a different problem. To see what it is, let us consider a particular version of this approach which has a venerable tradition in

[32]If one was attracted to Earman's definition, one might defend it in this way by saying that all directly perceivable structure supervenes on diffeomorphic structure.

[33]They do not *call* the functions defined in this way 'symmetries', but the functions that they define in this way are functions that they think play a role in the sorts of inferences outlined in Section 2. See my comments in footnote 29. The view is also implicit in Roberts [12].

the physics literature, a version which focuses on *relative positions*. For example, Bell wrote that 'in physics the only observations we must consider are position observations, if only the positions of instrument pointers', and that 'it is always positions that we are in the end concerned with'.[34] He does not say so explicitly, but it is charitable to read him as meaning *relative* and not absolute position. And recall the quote from Einstein that I originally set out on p. 16:

> All our space-time verifications invariably amount to a determination of space-time coincidences. If, for example, events consisted merely in the motion of material points, then ultimately nothing would be observable but the meeting of two or more of these points. Moreover, the results of our measurings are nothing but verifications of such meeting of the material points of our measuring instruments with other material points, coincidences between the hands of a clock and the points on a clock dial, the observed point-events happening at the same place at the same time.[35]

The idea here seems to be that every measurement outcome consists in a coincidence between two material events, which is a fact about their relative positions.

Both these authors seem to be gesturing at a view on which every directly perceivable property either is a property concerning relative positions, or at least supervenes on relative positions. If one was attracted to the idea, and if one was attempting to formulate an ontic definition of symmetry, it would therefore be tempting to say that F is a symmetry of L just in case F preserves the truth of L and also preserves all facts about relative positions. Since, on this view, relative positions play a privileged role in our epistemology, there would be nothing gerrymandered or *ad hoc* about the definition.

But the problem becomes apparent as soon as we ask for more detail about what "relative positions" really are. The obvious reading is that they are spatial distances between things, but in that case the definition fails immediately since the Lorenz transformations do not preserve spatial distances and yet are symmetries of STR (indeed, they are paradigm cases of symmetries!). So, alternatively, one might propose a more theoretical account of what one meant by relative positions: for example, one might say that relative positions are really space-time intervals (which are preserved by Lorenz transformations). Obviously the resulting definition would now allow Lorenz transformations to be symmetries, but it is now less obvious that relative positions—understood to be space-time intervals—really provide a

[34] Bell [1], p. 166.

[35] Einstein [3] p. 117, quoted in Earman [4] p. 187.

supervenience basis for all the directly perceivable properties. After all, they are insufficient to fix spatial distances, shapes and colors, but aren't these exactly the sorts of properties that we always thought *are* directly perceivable?

Of course, the reply will presumably be that STR has taught us that these sorts of properties are not directly perceivable after all, and indeed are not even real. And this is entirely accurate—indeed, I outlined the relevant "symmetry considerations" in Section 2: since these features vary under the symmetries of STR, we infer that they are undetectable (by the Symmetry-Reality Link) and unreal (by the Symmetry-Reality Link). Unfortunately, though, this reply leaves the advocate of the current approach in an unstable position. Remember, we complained about Earman's definition because it picked out a seemingly arbitrary bit of structure—diffeomorphic structure—which it required symmetries to preserve. As a result, the very definition of symmetry left diffeomorphic structure immune to the symmetry considerations, and we asked why that structure should be given such a priviledged role. In reply, we were told that structure is immune because it is directly perceivable. But now we are being told that we can use symmetry considerations to teach us about what sort of structure is directly perceivable after all!

So we are now in a position to offer a principled an argument against ontic definitions in general. Ontic definitions pick out a bit of physical structure and require symmetries to preserve it. And in order for the definition to avoid being *ad hoc* in the way that Earman's is, it must say why the designated bit of structure is immune to the symmetry considerations outlined in Section 2, whereby we infer that certain bits of structure are undetectable and unreal. But what the case of STR teaches us is that we have used the symmetry considerations on all sorts of properties that we might initially have thought were immune, such as the paradigm examples of directly perceivable properties! Thus, the problem is this: ontic definitions are committed to the existence of some sort of structure that is immune to the symmetry considerations, but given the case of STR we should be extremely suspicious as to whether any such structure exists.

Mentalistic Definitions

Ontic definitions made the mistake of thinking that what must be preserved by the symmetries of a law L (along with the truth of L itself) must be something out there in the world. The problem was that whatever sort of structure we chose would then be in principle immune from the symmetry considerations, and we have reason to be suspicious whether there is any such structure. But the problem dissolves once we stop looking outward to some structure in the world and instead look inwards to the contents of our minds. For even though spatial distances would be different in a Lorenz transformed world, my counterpart in such a world would nonetheless enjoy the very same kinds of experiential states as I do. Intuitively speaking, things would look and feel and taste and sound and smell exactly the same in the transformed world as it does in the original world. So perhaps we should require the symmetries of a set of laws to preserve (not only the truth of those laws but also) the kinds of experiential states enjoyed by subjects.

Any definition of symmetry along these lines is what I call a "mentalistic" definition, and it is this approach to symmetry that I want to defend. Now, there is not enough space to present or defend a particular mentalistic definition here, so I will present a definition scheme instead. The scheme will appeal to two place-holder notions, and while I will say a little about how such notions might be cashed out I will not defend any particular precisification.

The intuitive starting point for a mentalistic definition of symmetry is simple this: that the mental state I am currently enjoying is in some intuitive sense indiscriminable from the one I would be enjoying were everything, for example, uniformly boosted by some constant velocity. So any mentalistic definition will need to introduce a relation of *indiscriminability* between mental states. At a minimum, then, a function F on possible worlds will only count as a symmetry if it maps worlds W onto worlds F(W) in such a way that the mental states of agents in F(W) are indiscriminable from those of their counterparts in W. Now, while we have (I take it) at least some grip on an intuitive notion of indiscriminability, there are a number of ways to precisify this notion. So this is the first placeholder notion I will appeal to, and I will discuss some ways of precisifying it below.

But merely requiring that the mental states of agents in F(W) are indiscriminable from those of their counterparts in W is not strong enough to capture the intuitive idea behind mental definitions, for we have so far put no constraint on what F(W) is like in regions far away from any concious agents. The intuitive idea behind the current approach is that we want the entire *world* F(W) to be "indiscriminable" from W. But what is it for two worlds to be indiscriminable? One *could* try to say that two worlds are indiscriminable just in case they agree on some specified sort of structure, e.g. "directly perceivable" structure—but then the approach would collapse into the ontic approach rejected above. So, in keeping with the current mentalistic approach, let us understand the relation of indiscriminability between possible worlds in terms of the relation of indiscriminability between mental states. How this is to be done is far from obvious, but here is a rough gloss: worlds W1 and W2 are *indiscriminable* just in case the following condition holds: for any region R1 and direction D1 in W1, if there were an agent like us located in R1 and facing direction D1, the mental state she would enjoy would be indiscriminable from that which her counterpart would enjoy in W2.

Now, this definition is given in terms of a counterfactual and there are many questions about how exactly to understand it. So this definition is the second placeholder I will appeal to, and I will discuss some ways of precisifying it in a moment. For now, though, we can offer the mentalist definition scheme: a function F on possible worlds counts as a symmetry of a law L just in case (i) F necessarily preserves the truth of L, and (ii) F(W) is indiscriminable from W, for all W.

This approach avoids the problems that plagued our previous attempts to define symmetry. For one thing, it does not result in the Symmetry-Detection Link being prone to the sorts of counterexamples that plagued the attempt to define symmetry in purely set-theoretic terms. This is because, plausibly, the sorts of gerrymandered functions on worlds that would count as symmetries on those definitions and which falsified the Symmetry-Detection Link would not map worlds to indiscriminable worlds, and would therefore not count as symmetries according to the mentalistic approach. And unlike ontic definitions, it does not pick out some privileged bit of physical structure that symmetries are required to preserve by definition; so, unlike ontic definitions, it is not committed to the problematic claim that there exists some bit of physical structure that is immune from the symmetry considerations. Nonetheless, our mentalistic approach can nicely diagnose why ontic definitions might initially have seemed attractive. For we all hold a "folk theory of experience" telling us what sorts of experiences agents would come to have in different physical environments, and part of that folk theory tells us that if two environments agree on all facts about (say) the relative positions of things, then agents in those environments will enjoy indiscriminable mental states. So it is not surprising that, as a shorthand and in an attempt to keep a discussion tractable, theorists were tempted to give relative positions themselves a central role when thinking about symmetry. *Mutatis mutandis* for other bits of physical structure appealed to in ontic definitions, such as other "directly perceivable" properties. All these approaches were motivated by the entirely correct idea that there are systematic correlations between various bits of physical structure and our experiences—they just made the mistake of building that physical structure itself into the definition of symmetry. We might charitably see them as "approximations" to our mentalistic definition, convenient shorthands that suffice for the given job at hand even if they miss the underlying essence of symmetry.

So mentalistic definitions avoid the problems that beset other definitions while also diagnosing why those other definitions were initially attractive. But there is more to be said in favor of mentalistic definitions. Remember, we are looking for a definition of symmetry that not only makes the Symmetry-Detection Link true, but which also makes sense of the sort of argument run in the Galilean Two-Step. And mentalistic definitions seem to fit this bill well. To see why, suppose that S is a symmetry of a set of laws L, and suppose a given feature F varies under S. The Galilean Two-Step then attempts to show is that if the laws of L were true and complete, we would not be able to detect F. Now, when we rehearsed the argument in the last section, we reasoned in Step One about how physical devices would behave under various counterfactual initial conditions in which the values of F differed, on the basis of facts about the symmetries of L. And in Step Two we had to argue that it follows that F is undetectable. While Step Two was plausible in the case of NGT and absolute velocity discussed there, we noted that it actually needs to bridge a rather large gap. Indeed, we saw in the responses from Al and Bob that it is not in fact obvious why facts about how devices would behave under counterfactual initial conditions—which is all that was established in Step One—imply epistemic facts about what we can detect.

But with our mentalistic definition in hand, we avoid these problems. Instead of reasoning about how physical devices behave under counterfactual initial conditions that differ in the values of F, we use Step One to reason about *what sorts of experiences* we would enjoy under those counterfactual initial conditions—specifically, that the experiences we would have would be indiscriminable. The reasoning is simple. Again, suppose that S is a symmetry of some laws L, and suppose that a given feature F varies under S. If we let W be a world in which the laws of L are true and complete, then by supposition the values of F are different in W than in S(W). And because S is a symmetry L, it follows from our definition of symmetry that S(W) is indiscriminable from W: the experiences of any agents in S(W) are indistinguishable from their counterparts in W. Crucially, it also follows from S's being a symmetry that S(W) is a world in which the laws of L hold. So it follows that if the values of F were different from their values in W, in the way given by S, agents would (given the laws L) enjoy experiences indiscriminable from those enjoyed in W. That is Step One. In Step Two we then need to argue that F is undetectable, but this now follows easily: after all, it was established in Step One that it is physically impossible for the values of F to have any discriminable impact on our mental life, and it seems to follow that F is indeed undetectable.

In sum, our mentalistic definition makes good sense of the sort of reasoning rehearsed in the Galilean Two-Step, and (thereby) makes the Symmetry-Detection Link plausible—which is precisely what we were looking for in a definition of 'symmetry'. Moreover, our mentalistic approach also diagnoses why our original presentation of the Galilean Two-Step reasoned about how physical devices behave, rather than about what experiences we would have. For as we saw above, we all hold a "folk theory of experience" according to which if two agents are looking at bits of paper with the same pattern of ink on them, they will come to have indiscriminable experiences. Therefore, establishing that any physical device would produce the same pattern of ink on paper whether it was turned on at rest or in motion is a convenient short-hand for establishing what is (from the perspective of a mentalistic approach) the relevant result, namely that our experiences in a boosted would would, as a matter of physical necessity, be indiscriminable from our actual experiences.

Remember, the mentalistic definition I outlined is really a definition scheme since it appealed to two notions that I did not precisify, so let me finish by saying a few words about each notion. First, I appealed to the relation of indiscriminability between mental states, but I did not say anything about the conditions under which mental states are indiscriminable. Here are a few suggestions of such a condition:

> That they have identical phenomenologies
>
> That they have indistinguishable (though perhaps not identical) phenomenologies
>
> That they have identical perceptual contents
>
> That they have indistinguishable perceptual contents

No doubt there are many other possibilities too. Which notion of indiscriminability we should pick depends, I think, on some tricky issues in the philosophy of mind such as the relation between phenomenology and perceptual content, and also some tricky issues concerning the epistemology of perceptual judgement. But of course there is no way I can settle these issues here, which is why I leave the relation of indiscriminability as a placeholder.

The other placeholder was the counterfactual I used to define the relation of indiscriminability between possible worlds. Remember, I said that worlds W1 and W2 are *indiscriminable* just in case the following condition holds: for any region R1 and direction D1 in W1, if there were an agent like us located in R1 and facing direction D1, the mental state she would enjoy would be indiscriminable from that which her counterpart in W2 enjoy. This leaves open a lot of questions. What exactly makes an agent one that is "like us"? And what is being held fixed when determining what sorts of mental states an agent would have in that environment in W2: the laws of W1 or the laws of W2? These are extremely delicate questions that I will not be able to settle here. But I should emphasize that there *must* be some legitimate understanding of the counterfactual, on pain of our physical theories being untestable! To see this, note that physical theories describe the behavior of highly theoretical systems: electrons orbiting nuclei, wavefunctions undulating in various directions, and so on. To test such a theory, we must have some way of "interpreting" the theory or giving it "empirical content", as it is sometimes put in the literature. Plausibly, this requires determining what sorts of mental states an agent would come to have were they were coupled with the system in a certain way, for we can then test the theory by seeing whether we actually do come to have those experiences when coupled with such systems. Obviously, this requires that there be some fact of the matter as to what sort of mental state we *would* enjoy under those counterfactual conditions, else our actually coming to enjoy that mental state, when testing the theory, would provide no evidence for or against it. The job of clarifying this sort of counterfactual is therefore central to the problem of "interpreting" or giving "empirical content" to physical theories. And, of course, once that counterfactual is clarified, the counterfactual used in our definition of indiscriminability between worlds will *ipso facto* be clarified too. But again, this is an issue that I cannot hope to settle here, so for now I leave the counterfactual as a placeholder.

6 Conclusion

This completes my rational reconstruction of the role of symmetry in physical reasoning. The reconstruction starts by insisting on a mentalistic definition of symmetry, rather than an ontic or formal one. It then proceeds to use the Galilean Two-Step to argue that the Symmetry-Detection Link holds. And it then uses this link to justify our use of symmetry as a guide to reality and our use of symmetry as a guide to theory.

References

[1] Bell, J. S. 1993. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: CUP.

[2] Einstein, A. 1905. Zur Elektrodynamik bewegter Korper. *Annalen der Physik* 17: 891.

[3] Einstein, A. 1916. The Foundation of the General Theory of Relativity. Reprinted in English in *The Principle of Relativity*, 1952, edited by W. Perrett and G. B. Jeffrey. New York: Dover.

[4] Earman, J. 1989. *World Enough and Space Time*. Cambridge: MIT Press.

[5] Feynman, R. 1963. *Lectures on Physics: Volume 1*. Reading, MA: Addison-Wesley.

[6] Howard, D. 1999. Point Coincidences and Pointer Coincidences: Einstein on the Invariant Content of Space-Time Theories. In *The Expanding Worlds of General Relativity* pp. 463–500, edited by H. Goenner, J. Renn, J. Ritter and T. Sauer. The Center for Einstein Studies.

[7] Huggett, N. 1999. *Space from Zeno to Einstein*. Cambridge: MIT Press.

[8] Maudlin, T. 1994. *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics*. Oxford, UK: Blackwell.

[9] Maudlin, T. 1993. Buckets of Water and Waves of Space: Why Spacetime Is Probably A Substance. *Philosophy of Science* 60 (2): 183–203.

[10] North, J. 2009. The "Structure" of Physics: A Case Study. *Journal of Philosophy* 106: 57–88.

[11] North, J. Structure in Classical Mechanics. *Manuscript*.

[12] Roberts, J. 2008. A Puzzle About Laws, Symmetries and Measurable Quantities. *British Journal for the Philosophy of Science* 59: 143–68.

[13] Tipler, P. A. and G. Mosca. 2004. *Physics for Scientists and Engineers: Fifth Edition*. NY: W.H. Freeman and Company.

[14] Wigner, E. P. 1967. *Symmetries and Reflections*. Bloomington, IN: Indiana University Press.