On Surplus Structure Arguments

November 19, 2018

Abstract

Surplus structure arguments famously identify elements of a theory regarded as excess or superfluous. If there is an otherwise analogous theory that does without such elements, a surplus structure argument prompts adopting it over the one with those elements. Despite their prominence, the form, justification, and range of applicability of such arguments is disputed. I provide an account of these, following Dasgupta ([2016]) for the form, which makes plain the role of observables and observational equivalence. However, I diverge on the justification: instead of demanding that the symmetries of the theory relevant for surplus structure arguments be defined without recourse to any interpretation of those theories, I suggest that the process of identifying what is observable and its consequences for symmetries work in dialog. They settle through a reflective equilibrium that is responsible to new experiments, arguments, and examples. Besides better aligning with paradigmatic uses of the surplus structure argument, this position also has some broader consequences for scope of these arguments and the relationship between symmetry and interpretation more generally.

1 Introduction

The process of theorizing about the physical world is difficult. This is one reason why scientists and philosophers look to symmetries, which are held to have potential metaphysical and epistemological significance (Brading *et al.* [2017], §5). For instance, suppose one can exhibit that, according to a certain theory, a structure representing a putative feature of the world varies under application of symmetries of the theory—it is surplus structure (Redhead [1975], p. 88). 'Symmetries can be a potent guide for identifying superfluous theoretical structure' (Ismael and van Fraassen [2003], p. 371) as, in their presence, one has reason to believe that, ceteris paribus, the variants of this structure represent no real distinctions in the phenomena that theory represents.

But following this guide is no trivial matter. This is because excess, surplus, or superfluous structure in a theory does not typically announce itself. Proceeding semantically (rather than syntactically), one must first 'generate a set of models rich enough to embed the phenomena, [then] attempt to simplify those models by exposing and eliminating down excess structure' (Ismael and van Fraassen [2003], p. 390) that is idle in how the theory represents the phenomena. Classic examples include Leibniz's static and kinematic shift arguments against absolute space using the Principles of Sufficient Reason and Identity of Indiscernibles,[1] and

[1]There is a large literature on this type of argument. See, for instance, Hacking ([1975]), Belot, Belot ([2001, 2003]), Ismael and van Fraassen ([2003]), Dasgupta ([2016]), and

the use of U(1) gauge transformations to argue against absolute potentials in electromagnetism (Healey [2007]).

Although many authors recognize the validity of arguments roughly of this form, there is still some disagreement as to what that form is, and its own range of validity (Belot [2013]). Dasgupta ([2016]) has however recently brought new analytical focus to the structure of this argument and what it entails for the nature of symmetry thus invoked.[2] He argues that there are two ways in which the surplus structure argument involves epistemic considerations. First, they license inferences from the undetectability of surplus structure in a theory to an observationally equivalent theory, ceteris paribus, in which that structure does not represent any real feature of the world. Second, antecedent analysis of which worldly features are detectable through our faculties of perceptions, prior to and independent of the metaphysics of the theories to which it is applied, must justify what get to count as symmetries of observables.

I agree with the first but part ways with him on the second. In particular, while I agree that it is an Occamist norm to which the surplus structure argument appeals, I also point out several unmet obstacles to defining observational equivalence prior to any interpretation of the theories to which it is supposed to apply. Nevertheless, since we agree on the form of the surplus structure argument, his essay will be a touchstone for motivating my own account in section 4: roughly, that observability in and interpretation of theories proceed hand-in-hand through a process of reflective equilibrium. This account also fits better the historically paradigmatic uses of the surplus structure argument, and how scientists responded to cases in which what was previously thought to be surplus structure was not.

In the sequel, after developing some elementary formal tools for describing symmetry in section 2, I proceed to recount and critically evaluate Dasgupta's twofold epistemic proposal in section 3. I present my alternative, reflective equilibrium proposal for the relationship between interpretation and observation in section 4, including sketches of how it better

references cited therein.

[2]As I remark tangentially at some points (and especially in footnotes 7 and 8), Dasgupta's position is (despite his asseverations to the contrary) very similar to that of Ismael and van Fraassen ([2003]). Nevertheless his presentation deserves the touchstone honor (as I say in what follows) for it is much more perspicuous on the present items of discussion.

coheres with historical examples. I conclude in section 5 with some positive remarks (like those of Dasgupta ([2016])) on the scope of surplus structure arguments against the skepticism of Belot ([2013]): the arguments do have a generally valid form, once we are attentive to certain details about different types of symmetry and the role of isomorphism vis-á-vis representational capacities.

2 Symmetries of Models and Theories

Let $\mathcal{S}$ be the space of models (states) of a physical system characterizing some phenomenon of interest. They represent the various states of affairs in which the system could be, according to some theory. The states have various properties, and the values of those properties partition $\mathcal{S}$ into equivalence classes. For example, if X is a Boolean property defined for all models—i.e., it either obtains or does not for any particular state—then it induces an equivalence relation $\equiv_X$ on $\mathcal{S}$ whose two equivalence classes consist, respectively, of those models of $\mathcal{S}$ in which X obtains, and those in which it does not. If X is a quantitative property defined for all models—e.g., like mass, it takes on values in the non-negative real numbers—then $\equiv_X$ will induce as many equivalence classes on $\mathcal{S}$ as there are values of X that obtain in at least one model of $\mathcal{S}$. Observable properties, such as relative distances and times, although not distinguished formally from non-observable ones, would be examples. Note while many properties of interest are defined for all models, some are not: considering matter fields in Newtonian and Galilean spacetimes, only in the former do the fields have a well-defined state of absolute motion.

There may also be relevant relations (or other structures) defined on $\mathcal{S}$, too, besides the equivalence relations induced from properties defined thereon. Here I only focus on the case of a binary relations $\sim$ of observational indistinguishability. There may be more than one such relation, as they may be indexed to particular theoretical accounts of and technological capabilities for measurement. For the purposes of this essay, I shall assume that these relations are reflexive and symmetric, but (unlike an equivalence relation) not transitive.[3] In other

[3]These assumptions may need to be weakened in some examples (Tversky [1977], Fletcher [forthcoming]), but doing so doesn't materially affect the details of my arguments in sections

words, any model represents a state of affairs observationally indistinguishable to itself, and if one is observationally indistinguishable from another, the other is from the one. Sorites-like considerations illuminate they are not transitive: shades of a color may be to the typical naked eye observationally indistinguishable from those a bit darker and brighter, but transitivity would then imply that the blackest shade would be indistinguishable from the whitest.

Next, let $\mathcal{T}$ be the set of bijective functions on $\mathcal{S}$, called *transformations of* $\mathcal{S}$. Some of the transformations *preserve* properties, relations, or other structures on $\mathcal{S}$, and these will be called symmetries. But it is important to distinguish a symmetry by which structures it preserves. Given an equivalence relation $\equiv$ on $\mathcal{S}$, T is an $\equiv$-symmetry of $\mathcal{S}$ just when for any $s \in \mathcal{S}$, $s \equiv T(s)$. An $\equiv$-symmetry preserves the valuation of the property inducing the equivalence relation $\equiv$. Similarly, given any relation $\sim$ of observational indistinguishability on $\mathcal{S}$, T is a $\sim$-symmetry of $\mathcal{S}$ just when for any $s \in \mathcal{S}$, $s \sim T(s)$. A $\sim$-symmetry preserves enough features of a model that the transformed model is observationally indistinguishable from the un-transformed model.

When all the models in $\mathcal{S}$ have certain features in common, sometimes one can describe a transformation on these features that *induces* a transformation on the models, which one can then consider as a candidate symmetry. One way this can happen is if there is a property assignment that is well-defined for all models. In the Boolean case, a mapping that switches the valuations 'true' and 'false' will induce a map on the models that send any model to one that is the same but with the valuation of that property switched. In the quantitative case, a bijection—really, an isomorphism—of the valuation space induces a map on the models in the same way. For instance, in Newtonian spacetime models, this would include any velocity shift of the absolute rest frame. Note, however, that there must be for each model in $\mathcal{S}$ a well-defined model that is the image of the putatively induced map. Shifting the velocity of the absolute rest frame in Galilean spacetime is not well-defined.

Another (compatible) way to induce a transformation on the models arises when the models assign properties to the same set of objects. A permutation of those objects then induces a transformation on the models. Often, one desires such a permutation to also preserve any

3 and 4.

special structure these individuals may have. For instance, consider the collection of models that use a common fixed smooth manifold to represent spacetime: each of its models assigns various properties to individual events (points) and collections thereof. Diffeomorphisms of the smooth manifold permute the events while preserving its smooth structure. Such maps induce pushforward maps that are then candidates for being isometries, or some other type of structure-preserving mapping (symmetry).

3 Dasgupta's Epistemic Proposal

3.1 The Account

Having presented some elementary formal apparatus to make symmetry reasoning with physical theories more precise, I can proceed to form of the surplus structure argument. Dasgupta ([2016], p. 853) in particular reconstructs it as follows:

(1) Laws L are the complete laws of motion governing our world.

(2) Feature X is variant in L.

(3) Therefore, X is undetectable (from (1) and (2)).

(C) Therefore, X is not real (from (3) and the Occamist norm that we dispense with undetectable structure).

Some explication of Dasgupta's terms is in order. First, for present purposes one may think of the laws L as selecting a subset of the state space $\mathcal{S}$ of the universe (cf. 'our world') that count as dynamically possible, or more generally and charitably, as nomologically possible.[4] (Thus the state space may be said to represent mere metaphysical or kinematic possibility.) Because L partitions the states into those satisfying the laws and those not, it induces an equivalence relation $\equiv_L$ on $\mathcal{S}$. A feature X is just a property, and the values that it takes on different elements $\mathcal{S}$ also partition them into equivalence classes corresponding to the induced equivalence relation $\equiv_X$. For X to be variant in the laws is to say that there is an $\equiv_L$-symmetry* transformation on $\mathcal{S}$ that is not an $\equiv_X$-symmetry* transformation on $\mathcal{S}$. Here, a *symmetry** is a

[4]Not all (even physical) laws need be laws of motion. This makes no difference in what follows, however.

special type of symmetry in the sense I have defined in section 2, one that is induced from a transformation on properties or individuals,[5] and is also a symmetry of *all observational/detectable features* (Dasgupta [2016], §6.4), understood as equivalence relations in the same way as X. The intermediate conclusion, (3), follows because if X were detectable, it would be preserved by a symmetry*, but this would be just to deny the premise (2).

I will return to Dasgupta's account of symmetry* shortly, but first I complete the exegesis of his account of the form of the symmetry argument against surplus structure. In the first place, 'If we can show that the feature [X] is undetectable, then we will have shown that we do not (and cannot) have empirical evidence in the form of observations or measurements of it' (Dasgupta [2016], p. 854).[6] Then one can infer to a theory with the same laws L but in which X does not mark out any real differences, provided that 'we have the alternative theory in hand and have shown that all else is equal' (Dasgupta [2016], p. 854), in particular, that 'the theories that dispense with the feature [X] do not have [additional] other epistemic vices' (Dasgupta [2016], p. 854) such as inelegance or complexity.[7] Thus the inference to (C) is only based on ceteris paribus considerations. But there are clear cases where it does apply: without Galilean spacetime known to him, Newton was right not to reject the reality of absolute space in the face of Leibniz's static and kinematic shift arguments, but now that this option is known, one should so reject it.

[5]I've expressed this clause as a disjunction since Dasgupta ([2016], p. 858n27) doesn't seem to distinguish between transformations on properties from those on individuals; what matters to him is that such transformations are 'generated by a recipe' (Dasgupta [2016], p. 858).

[6]Actually, Dasgupta ([2016], p. 874) restricts attention to undetectable properties that are also not defined in terms of detectable properties, because one does have empirical reason to believe those are real—namely, whatever reason one has for believing in the reality of the detectable properties in terms of which they are defined. Technically, this only solves the problem for properties that are explicitly so-defined, so an extension to implicitly defined properties would make sense.

[7]See also Ismael and van Fraassen ([2003], p. 390–1) for a similar observation on the scope of this sort of symmetry argument against surplus structure.

It should be clear that symmetry* cannot be weakened to symmetry (sans asterisk), which is only formally defined, for there would then be no necessary connection between L-symmetry and undetectability, rendering the inference to the intermediate conclusion (3) invalid. Moreover, Dasgupta ([2016], p. 867, emph. add.) insists that

> observational equivalence must be defined in epistemic terms that do not depend on the underlying metaphysics, such that we can be in a position to know whether two structures [i.e., models] are observationally equivalent *prior to knowing anything* (via symmetry-to-reality reasoning [i.e., the type of argument form under discussion]) *about the metaphysics of our world*.

This is because otherwise one is led to an inferential circularity—a circularity of justification—in the symmetry argument against (the reality of) surplus structure, as

> we often use premises about symmetries[*] in order to work out which physical features fix the [observational] data, so we cannot at the same time define symmetries[*] to be those operations that preserve features that fix the [observational] data (Dasgupta [2016], p. 865)

if we wish to maintain the general validity of this argument (if only on the ceteris paribus conditions discussed above) . Consequently, these arguments broach 'considerations that reach into the philosophy of perception and mind' (Dasgupta [2016], p. 875).[8]

[8]Pace Dasgupta ([2016], p. 867), his epistemic definition of (what I am here calling) symmetry* is essentially the same as that of Ismael and van Fraassen ([2003], p. 379, emph. orig.), namely, that the symmetries (unstarred!) at play are those '*that also preserve all qualitative features of every model*' of the theory, i.e., are L-symmetries, where qualitative features of a model are that those 'which characterize that [model], and are directly accessible to us through perception' (Ismael and van Fraassen [2003], p. 375). Dasgupta takes Ismael and van Fraassen to be defining qualitative features from symmetries, but they are quite explicit that they are not doing so: their definition of qualitative features (Ismael and van Fraassen [2003], p. 375) comes before their definition of symmetry (Ismael and van Fraassen [2003], p. 378–9) and does not invoke it.

It would be a surprising conclusion that implicates more commonplace understandings of symmetry, including essentially all those applied by physicists, as being insufficiently attentive to the nature of perception. But what positive proposal does Dasgupta make? He associates with each model $s \in \mathcal{S}$ and each *context* i within the model—this is not defined precisely, but should be something like a frame of reference—that determines a class of observation sentences deemed 'correct,' suggesting that 'one's grasp of observations sentences . . . allows one to determine their correctness-conditions in each kind of world, including what kinds of indices correctness relativizes to' (Dasgupta [2016], p. 870). Any transformation on $\mathcal{S}$ induces a transformation on the index i, which in turn induces a transformation on the set of observation sentences. If this induced transformation on the observation sentences is the identity for *all* models and *all* indices, it is said to preserve the observation sentences. Then an $\equiv_L$-symmetry* is one that preserves not only L but also the observation sentences. Crucially, once one grasps the contextualized correctness-conditions 'it is then an *a priori* matter whether the transformation is a symmetry of the law' (Dasgupta [2016], p. 870), one that does not presuppose any underlying metaphysics.

3.2 The Objections

However, there are a number of unaddressed problems with this suggestion that make the class of symmetries* too *small*. I'll set aside concerns about underspecification of the indexing proposal, and the usual objections that observation sentences are theory-laden (Bogen [2017]) and so do in fact presuppose metaphysical claims on which Dasgupta wishes to remain agnostic.[9] Instead I'll focus on the lack of constraints on both what the observation sentences and the models under consideration can be. Consider a putative observation sentence, 'My only firecracker went off at event p.' Now, any nontrivial spacetime transformation T, such as

[9]There's the additional, related issue that the laws L may well presuppose spacetime structure, so that determining whether a transformation is an $\equiv_L$-symmetry* also presupposes settling certain of the claims on which Dasgupta again wishes to be agnostic. He seems to acknowledge this point (Dasgupta [2016], p. 848n14) but sees it as a 'distraction' from his main argument.

a diffeomorphism on the spacetime smooth base manifold, will map some spacetime point to another not identical to it, e.g., $p \mapsto q \neq p$. So if the sentence is acceptable for s, it will not be so for $T(s)$. No spacetime transformation that does not act as the identity on p can thus be a symmetry*, regardless of the laws, if such sentences can be countenanced as acceptable observation sentences.

Now, it is precisely the moral of the hole argument in general relativity (on one reading, at least) that such transformations induce no change because they induce no observable change, as Dasgupta ([2016], p. 840–1) himself describes. But that is exactly the sort of argument of which he cannot avail himself. It doesn't seem *metaphysically* impossible that there could be beings who could observe the essence of spacetime events themselves. How can such observation sentences therefore be ruled out as being acceptable on epistemological grounds, without already employing some assumptions about what the terms in the sentences represent? Without ruling them out, it becomes difficult to find many of the symmetries* we antecedently supposed were present in theories we already use.

It isn't clear to me that they can be ruled out a priori, but suppose they could. There is still the problem that even for observation sentences that don't refer to seemingly unobservable features, the class of *models* will be too wide to allow for any nontrivial symmetries*. Because a symmetry* must preserve observation sentences for *all* models, a very wide class of models makes it more difficult for a transformation to be a symmetry*.

I have in mind in particular two classes that will be too wide. One class consists of models only on some of which certain properties are well defined, or which do note predicate features of the same individuals. The result is that it is even harder to find transformations—candidates for symmetries*—that are induced from transformations on properties of models or on the individuals these models describe. Such classes are not entirely chimerical, for they can result from juxtaposing the models of a theory with those that are the *result* of expurging surplus structure therefrom! To take two familiar examples with properties: static shifts that move the center of the world are well-defined for Aristotelian spacetime, but not for Newtonian ones; kinematic shifts that adjust absolute velocities uniformly are well-defined for Newtonian spacetime, but not for Galilean ones. And two slightly more esoteric examples that seem

nevertheless to be physically or metaphysically possible, in principle: a parity transformation that inverts left and right is well-defined on orientable spacetimes but not on non-orientable ones; spatial translations and rotations may not be defined for spacetimes with unusual topologies. Such transformations cannot count as symmetries*, even of the models we would otherwise regard as having a symmetry, because the transformations do not (indeed, *cannot*) act on the whole set of models.

For each putative symmetry*, I reckon that one can find models to add to $\mathcal{S}$ for which it is ill-defined—the models defined by the quotient of models under the putative symmetry*. One might argue that such putative symmetries* should be extended to act as the identity on these new models, but that would violate Dasgupta's resolve to define transformations on models in terms of transformations on their properties or individuals. With no property of a particular sort to be had for a certain model, the transformation thereby generated cannot have that model in its domain.

The other class of models I have in mind consists of those whose correct observation sentences are distinct from any other (even with index structure accounted for). These describe states that may be observationally quite like our own but whose apparent symmetries are in fact only approximate. This is entirely compatible with our observational evidence, and so it is entirely possible that 'perhaps one would find that none of the symmetries [that our theories attribute to Nature] is really exact if our observational/experimental capabilities were sufficiently improved' (Sundermeyer 2014, p. 12). By definition, then, no transformation on $\mathcal{S}$ other than the identity will induce a transformation on their observation sentences that preserves them for *every* context i,[10] for even if our present coarse means of observation renders the states indistinguishable, there will be some possible finer means that will resolve them. It will not help to modify Dasgupta's definition of symmetry* to range only over the present context i, for observational indistinguishability is not an equivalence relation as is also required by the definition: if states differ in continuous degrees, then the transitive closure of observationally indistinguishable states would include observationally distinguishable states.

[10]In the terminology of Caulton ([2015]) that I discuss below, these are models from a theory where identity is the only analytic symmetry.

Thus, when models without any symmetries are present in $\mathcal{S}$, there will be no symmetries*. To be clear, the argument is not that there is something wrong with reasoning that 'which physical features we should believe are real depends on which systems are observationally equivalent' (Dasgupta 2016, p. 872)—on this point I essentially agree with Dasgupta. Rather, the point is that the existence of states whose observation sentences are subtly distinct from those of any other states confutes the proposed definition of symmetry* that *would* account for its use in surplus structure arguments.

Finally, one could also raise objections to the descriptive accuracy of Dasgupta's epistemology-first approach to paradigms of the surplus structure argument from the history of physics. I will discuss these, however, in the course of extolling the virtues of my own account of the surplus structure argument in section 4.

4 A More Naturalistic and Historically Descriptive Proposal

4.1 The Account

Perhaps there is some other way of amending Dasgupta's proposal, but I shall not pursue it. For a methodological naturalist such as myself, the project of a priori epistemology and philosophy of perception is on just as dubious grounds as a priori metaphysics (Papineau [2016], §2). But regardless of one's methodological proclivities, and the status of my above objections, there is a different positive approach worth articulating. Instead of insisting on a foundationalist epistemology for the role of symmetries in surplus structure arguments—working out, in other words, exactly what is observable before using those facts to determine one's metaphysics—one can take a more coherentist approach and allow some metaphysics to influence what states and observations are possible (Olsson [2017]). For, my objections to Dasgupta's proposal above all turned on how the foundational epistemological purist has their hands tied behind their back in trying to constrain possible states and observation statements. Of course, it's important that any circles of justification in such an approach are not wound too tightly, but this does not present a significant problem: much of the metaphysics that one needs does not (pace Dasgupta ([2016], p. 864)) arise from such arguments.

The details employ a simple application of reflective equilibrium, inspired by a proposal for the interpretation of theories with symmetry by Caulton ([2015]).[11] One first proposes a plausible and empirically adequate (or nearly enough so) theory of a system of interest, which includes states, the observable properties thereof, and how these formal objects represent physical objects and properties. That theory's specification of its observable properties induces, for each one, an equivalence relation on its collection of states according to sameness of observable property. This in turn determines some symmetries, divided into the *analytic* and *synthetic*. What's important for present purposes are the analytic symmetries, which are those that preserve all the observable property equivalence relations.[12]

Second, one tests both that the class of observable properties is wide enough to account for the plethora of experience within the theory's scope, and that it is narrow enough that it doesn't include properties whose values don't bear on the empirical (or other, e.g., explanatory) value of the theory. This can include the application of surplus structure arguments for the function of narrowing, but it can also include empirical experimentation for the function of widening and of changing the possible states under consideration. (I'll give some examples below.) Eliminating states or observable properties can introduce more analytic symmetries, while introducing new states or observables can eliminate some analytic symmetries.

Third, one decides whether, ceteris paribus, one should identify or make synonymous (i.e., regard as equivalent) the remaining states related by an analytic symmetry, thereby eliminating the properties that those states did not have in common. Caulton ([2015], pp, 156,

[11]But there are differences. Caulton's proposal is not itself of a coherentist flavor nor does he advocate for reflective equilibrium; conversely, he asserts that his process leads to a unique interpretation, while I make no such claim.

[12]Thus analytic symmetries are close to what Ismael and van Fraassen ([2003], p. 380) call *trivial* symmetries, but they insist that the observable properties should be understood as those they call *qualitative*, which are those 'that can characterize a situation, distinguishable by even a gross discrimination of colour, texture, smell, and so on' (Ismael and van Fraassen [2003], p. 376). By contrast, Caulton ([2015], p. 150) insists that 'which symmetries count as analytic and which count as synthetic will hand on the details of this representation relation, and *vice versa*'. On this issue my proposal aligns with Caulton's.

160) recommends quotienting the models by the analytic symmetry equivalence classes (insofar as this does not reduce empirical adequacy), but the quotient operation does not always obviously yield a theory with practically feasible predictions or explanations of phenomena, as is arguably the case for holonomy interpretation of the Aharanov-Bohm effect Healey ([2007], pp. 120–2).[13] However, Dewar ([2017]) has recently pointed out that this is not the only option: instead of modifying the models, one can modify their isomorphisms, regarding analytic symmetries as isomorphisms. This works because isomorphic models have the same representational capacities: they may represent the same physical states equally well (Fletcher [2018]).

Which technique—quotienting models or adding isomorphisms—one choses depends on balancing the interpretative and explanatory considerations at hand. For example, insofar as one is less sure that a certain symmetry is exact, rather than approximate and liable to considered an accident from the vantage of future theory, one might lean towards adding isomorphisms, as it will be easier to 'undo' in future developments if it turns out to be so. But if one is sure, and doing so does not appreciably reduce the explanatory resources of the theory. It may even be advisable to take a pluralist stance towards these options, recognizing that they depend on weighing various factors that may not receive intersubjective agreement in the relevant expect community. Indeed, in the fourth step one returns to step two as new theoretical arguments and experimental facts about observables and their representation arises.

[13]Caulton ([2015], pp. 154) acknowledges that his proposal does not settle matters of interpretation that are hyperintensional, e.g., explanatory features of theory, but suggests that this is less relevant because the explanatory goals of physicists are distinct from the interpretive goals of philosophers (Caulton [2015], p. 161). I am not not so sure why ontology, interpretation, and explanation can be so cleanly separated. But in any case, as Dewar ([2017], pp. 12–14) points out, quotienting does seem to *eliminate* certain explanatory resources, such as those for showing *why* invariant quantities satisfy certain constraints, on which Caulton's proposal is not neutral. The point is not that one should never quotient, but rather that there are more constraints and considerations than empirical adequacy at hand.

4.2 The Objections, Met

Despite how philosophers may present the individual steps and arguments of this process in potted histories of, e.g., the transition from Newtonian to Galilean spacetime, each is usually a significant intellectual accomplishment. Ismael and van Fraassen ([2003], p. 372) do acknowledge this for steps two and three: 'A theory is usually around long before we know whether it contains unmeasurable quantities, and which quantities those might be. The process of identifying unmeasurable quantities, i.e. of smoking them out of their hidden places in the theoretical apparatus, is long, hard, and highly non-trivial.' But it is also so for step four. That's to say, it was significant both for Leibniz to develop his shift arguments, based on the identity of indiscernibles, *against* absolute space, time, and motion, and for Newton to develop his bucket and globes arguments *for* some sort of absolute time and motion. The former deploys a surplus structure argument (steps two and three), while the latter inserts a new argument that leads one (through step four) to re-evaluate the ceteris paribus clauses of that argument. I thus contend my proposal better aligns with scientific practice and the actual use of surplus structure arguments deemed successful through history. Because the process of reflection can accommodate new evidence and arguments, it also better accommodates how our judgments about what is observable and what symmetries are exact and what others merely approximate change over time, just as they have in fact. Accordingly, it does not require any assumptions about the *completeness* of the laws invoked, only their empirical adequacy for a certain domain of application (insofar as even this can be achieved), in contrast to Dasgupta's reconstruction.

It will help to illustrate this in the course of responding to a potential objection from Dasgupta ([2016], pp. 864–5). The objection reiterates the concern that 'we discover which physical features fix the data by engaging in symmetry reasoning' (Dasgupta [2016], p. 864), so that any determination of observables prior to determination of the (analytic) symmetries would preclude certain of these symmetries, e.g., ones we have actually adopted in the transition from pre-relativistic spacetime theories to the special theory of relativity.[14] This is to

[14]Dasgupta ([2016], p. 864) offers the concrete example of spatial distance: 'it is crucial that this is not counted as something that fixes the data, else Lorentz transformations will not

implicitly deny, in other words, the possibility of the fourth step in the iterative process of reflective equilibrium.

But the most cursory examination of historical and experimental examples involving symmetry reasoning about observational equivalence reveals a quite different conclusion. Any theory of empirical phenomena must interface with models of experiments, for which the theory's symmetries make predictions on the possible existence or non-existence of putative observables. For example,

> When we say [that there is] right-left symmetry, we imply that it is impossible to observe an absolute difference between right and left. . . . Indeed, all symmetries are based on the assumption that it is impossible to observe certain basic quantities, which we shall call 'nonobservables.' Conversely whenever a nonobservable becomes an observable, we have a symmetry violation. (Lee [1988], p. 6)

In other words, any (analytic) symmetry or lack thereof implies constraints or freedoms, respectively, in the phenomena to be observed. In the case to which Lee alludes, before 1956 physicists commonly supposed parity symmetry to hold. This predicts no observable absolute difference between right and left as predicted in physical theory, whereas in 1956 C. S. Wu and her colleagues discovered such a difference in beta decay experiments of cobalt-60 (Lee [1988], pp. 4, 11)!

---

count as symmetries of the STR!' However, the example is flawed: (proper) spatial distances do not vary under Lorentz transformations. Perhaps Dasgupta had in mind the phenomenon of length contraction, but that is an effect arising from coordinate re-descriptions (passive transformations) of the same events, which he denies is the subject of his analysis (Dasgupta [2016], p. 839n2). Alternatively, he could have had in mind two situations in which an object and its counterpart are in different states of motion relative to an observer (and its counterpart): the observer would indeed judge them to be of different sizes, for reasons analogous to those for length contraction. However, these two situations are *not* related by a Lorentz transformation—or any symmetry of Minkowski spacetime—at all, as only the object, not the observer, is boosted in one situation respect to the other.

Similarly, the relationisms of Cartesian and Leibnizian physics eliminated any observable absolute motions, but this then left a gap in explaining the inertial effects in Newton's bucket and globes thought experiments. And in the transition to special relativity, the lack of full spacetime Poincaré symmetry made the non-observability of the aether wind puzzling. The ad hoc FitzGerald contraction hypothesis notwithstanding, the transition to special relativity finally explained and settled the issue satisfactorily (Janssen [2002]). Finally, while assuming Poincaré symmetry in special relativity entails the existences of nonobservables—properties invariant under their transformations—this did not somehow preclude the development of general relativity, in which Poincaré symmetry does not hold. This list could be widened considerably.

5 The Scope of Surplus Structure Arguments

Although I agree with Dasgupta about the form of the surplus structure argument—in particular, that it turns on the nonobservable status of the structure to be marked as superfluous—I differ from him in the justification that we should (and do) give to which properties are observable. Rather than try to determine what is observable prior to any theory, which he calls an 'epistemic' definition of symmetry, I suggested instead a process of conjecture, argument, experiment, and reflective equilibrium that also better captures the historical vicissitudes of observability. In this last section I will draw out one further consequence of this position.

Dasgupta ([2016], p. 876) describes how his account renews optimism about the general validity of surplus structure arguments in the face of the skepticism of Belot ([2013]), which he attributes to his 'epistemic' definition of symmetry. However, the problems with Dasgupta's positive account for allowing for enough symmetries* and its misalignment with historical exemplars of the surplus structure argument, discussed in sections 3 and 4, respectively, would give the skeptic further pause. But because my account provides a different account of the justification for what counts as observable while sharing the same form for the surplus structure argument, it is in a better position to renew optimism about that argument's general validity.

To show this requires revealing in more detail the grounds for Belot's pessimism and how my account can be responsive to it. Belot ([2013], p. 319) begins with the following two principles that he see implicit in philosophical commentary on symmetries:

D1 The symmetries of a classical theory are those transformations that map solution of the theory's equation of motion to solutions of the theory's equation of motion.

D2 Two solutions of a classical theory's equation of motion are related by a symmetry if and only if they are physically equivalent, in the sense that they are equally well- or ill-suited to represent any particular physical situation.

D1 is intended as a formal definition of symmetry, while D2 is intended to connect that notion of symmetry with the interpretation of a physical theory. Combined, they would yield some way in which formalism and interpretation constrain each other. But literally understood, they imply immediately that any two models of a classical theory have the same *representational capacities*, i.e., they 'are equally well- or ill-suited to represent any particular physical situation.' Since most theories aim to describe more than singular phenomena, this is an unwelcome conclusion.

The significance of this for the surplus structure argument is that, at least in special cases, that argument may seem to involve commitments to D1 and D2, or principles near enough thereto (Belot [2013], p. 319n3). Take the example of a kinematic shift in a Galilean spacetime: it maps models to models, and does not alter spacetime structure, which determines the representational capacities of a Galilean spacetime model. On the other hand, a kinematic shift applied to a Newtonian spacetime shifts the absolute velocities of all matter and fields represented therein. Because of this mismatch between its symmetries and its representational capacities, Newtonian spacetime might be said to be postulating excess spacetime structure. But in light of the unwelcome conclusion above, this cannot be a generally valid argument form, and Belot ([2013], p. 333) places most of blame on D1 as a definition of symmetry: 'I leave it as a challenge to the reader to identify a general and interesting formal notion of symmetry [i.e., a rendering of D1] that renders D2 true.'

Under (my reading of) Dasgupta's reconstruction, however, the surplus structure argument proceeds rather differently. In the first place, I believe it deserves emphasizing more than those

writing before me that *symmetries come in many types* according to which equivalence relation (or otherwise) they preserve. It is important to distinguish them according to what exactly they preserve, let one fall into fallacies of equivocation. And that equivocation can leads to the unwelcome conclusion above.

The symmetries involved in the surplus structure argument are symmetries*: not just symmetries of the laws ($\equiv_L$-symmetries), which is what is specified in D1, but also symmetries of the (exact) observable properties of the models involved. The inference to deeming certain features superfluous runs not through a thesis like D2 about representational capacities, but rather through an Occamist norm about minimizing ontological commitments, ceteris pariubs. Rather than using something like D2 as a premise in the argument, the surplus structure argument's conclusion motivates adopting a theory in which the symmetries* relate models that are closer to being physically equivalent.

However, this is not to say, exactly, that D2 is the conclusion or a regulative ideal of the surplus structure argument, even with the 'symmetry' mentioned in D2 replaced by 'symmetry*'. The reason is that the biconditional does not hold for the same notion of 'symmetry' in both directions. Two models of a theory may be related by a symmetry* but not have the same representational capacities if the features variant between the models are not observational. Of course, applying a surplus structure argument to a theory with such models may motivate adopting *instead* a theory in which those models' respective counterparts *do* have the same representational capacities, but this is besides the point when it comes to the interpretation of the theory at hand. Being related by a symmetry* thus provides only a necessary condition for physical equivalence:

D2* Two models of a theory are related by a symmetry* if they have the same representational capacities (i.e., are physically equivalent).

For, if two models have the same representational capacities, they must be capable of representing in particular the same observable phenomena, so they may be related by some transformation that maps the one into the other.

Is there a sufficient symmetry-like condition for physical equivalence? I shall adopt the suggestion of Fletcher ([2018]), that there is:

D2' Two models of a theory are related by an isomorphism only if they have the same representational capacities (i.e., are physically equivalent).

In contrast with symmetry*, which is not an entirely formal notion due to its reference to observable equivalence, isomorphism is a formal feature of a category of mathematical models. Isomorphisms are maps between models that witness an equivalence between those models, as elements of that category. For example, two smooth manifolds are isomophic (diffeomorphic) when, as smooth manifolds, they are equivalent: among the mathematical properties of smooth manifolds, they have the same. So, using models from a particular mathematical category *as* members of that category to represent physical states of affairs entails that they have the same representational capacities, for otherwise they would not be models *as* members of that category.

In contrast with symmetry*, Fletcher ([2018]) has argued in particular that isomorphism is not a necessary condition for physical equivalence. The very existence of theories with surplus structure is proof of that, for that variant structure renders the models non-isomorphic even if it does no representational work. Even if such theories are deficient in some way, leading us to change the category by adding isomorphisms as Dewar ([2017]) suggests, it does not mean that they cannot be interpreted, and in any case, the ceteris paribus condition in that argument warns us that other (e.g., explanatory or pragmatic) considerations may void the attribution of superfluousness.

Although Belot's pessimism arose from fixing D2 and searching for a plausible analog of D1, I have suggested that D2 be replaced by two different principles that provide distinct conditions (one necessary, D2*, and the other sufficient, D2',) for physical equivalence. He has speculated along these lines: 'Of course, some interesting weaker relative of D2 might be true' Belot ([2013], p. 334n56). But he did not speculate that the analog of D1 that played a role in the surplus structure argument would not contain a formally defined notion of symmetry at all. Rightfully, it should involve a complicated interplay between theory, experiment, and representation, balanced continually through reflective equilibrium (as described in section 4). As Dasgupta ([2016], pp. 874, 876) warned, there seems to be no purely formal notion of symmetry that preserves the validity of the argument. Belot's

pessimism is warranted only when one demands that there be some such notion.

References

Belot, G. [2001]: 'The principle of sufficient reason', *The Journal of Philosophy*, **98**(2), pp. 55–74.

Belot, G. [2003]: 'Notes on symmetry', in K. Brading and E. Castellani (*eds*), *Symmetries in Physics: Philosophical Reflections*, Cambridge: Cambridge University Press, pp. 393–412.

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

Bogen, J. [2017]: 'Theory and Observation in Science', in E. N. Zalta (*ed.*), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, summer 2017 edition.

Brading, K., Castellani, E. and Teh, N. [2017]: 'Symmetry and Symmetry Breaking', in E. N. Zalta (*ed.*), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, winter 2017 edition.

Caulton, A. [2015]: 'The role of symmetry in the interpretation of physical theories', *Studies in History and Philosophy of Modern Physics*, **52**, pp. 153–162.

Dasgupta, S. [2016]: 'Symmetry as an Epistemic Notion (Twice Over)', *British Journal for the Philosophy of Science*, **67**, pp. 837–878.

Dewar, N. [2017]: 'Sophistication about Symmetries', *The British Journal for the Philosophy of Science*, p. axx021.
<http://dx.doi.org/10.1093/bjps/axx021>

Fletcher, S. C. [2018]: 'On Representational Capacities, with an Application to General Relativity', *Foundations of Physics*.
<https://doi.org/10.1007/s10701-018-0208-6>

Fletcher, S. C. [forthcoming]: 'Similarity Structure on Scientific Theories', in B. Skowron (*ed.*), *Topological Philosophy*, de Gruyter.

Hacking, I. [1975]: 'The identity of indiscernibles', *The Journal of Philosophy*, **72**(9), pp. 249–256.

Healey, R. [2007]: *Gauging What's Real: The Conceptual Foundations of Contemporary Gauge Theories*, Oxford: Oxford University Press.

Ismael, J. and van Fraassen, B. C. [2003]: 'Symmetry as a guide to superfluous structure', in K. Brading and E. Castellani (*eds*), *Symmetries in Physics: Philosophical Reflections*, Cambridge: Cambridge University Press, pp. 371–392.

Janssen, M. [2002]: 'Reconsidering a Scientific Revolution: The Case of Einstein versus Lorentz', *Physics in Perspective*, **4**, pp. 421–446.

Lee, T. D. [1988]: *Symmetries, Asymmetries, and the World of Particles*, Seattle: University of Washington Press.

Olsson, E. [2017]: 'Coherentist Theories of Epistemic Justification', in E. N. Zalta (*ed.*), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, spring 2017 edition.

Papineau, D. [2016]: 'Naturalism', in E. N. Zalta (*ed.*), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, winter 2016 edition.

Redhead, M. L. G. [1975]: 'Symmetry in Intertheory Relations', *Synthese*, **32**(1-2), pp. 77–112.

Sundermeyer, K. [2014]: *Symmetries in Fundamental Physics*, Cham: Springer, 2nd edition.

Tversky, A. [1977]: 'Features of Similarity', *Psychological Review*, **84**(4), pp. 327–352.