**In *The Philosophy of Philip Kitcher* ed. Mark Couch and Jessica Pfeifer, Oxford University Press, 2016.**

**Note: This version differs in minor ways from the published version. Please use the published version for citation purposes.**

**Unificationism, Explanatory Internalism, and Autonomy**

**1. Introduction**

Philip Kitcher has made an extraordinary number of distinguished contributions to philosophy of science and to many other areas of philosophy. Here I focus on just one of Kitcher’s projects in philosophy of science, although one that has been very influential—his development of a “unificationist” account of explanation. Although I will engage critically with some of Philip’s ideas on this subject, I also want to roam more widely, taking up some broader issues about the role of unification in explanation, the contrast between “internalist” and “externalist” approaches to explanation (in the sense of Kim, 1994), and claims about the autonomy of the special sciences and how these interact with ideas about unification. I will explore these issues against the background of the interventionist account of explanation I have defended elsewhere (Woodward, 2003). The reader should thus think of what follows as an investigation of some Kitcherian themes regarding explanation, rather than as a critical study that focuses just on Kitcher’s work.

The remainder of this essay is organized as follows. I begin (Section 2) with a brief summary of Kitcher’s ideas about the role of unification in explanation. I then (Section 3) turn to an overview of some of the principal claims for which I will argue, in the hope that this will help to guide the reader through what follows. Section 4 takes up some issues regarding internalism and Section 5 defends the importance of incorporating internalist as well as externalist considerations in models of explanation. Section 6 and 7 distinguishes two kinds of explanatory projects connected to unification.

**2. Kitcher on Unification and Explanation**

Kitcher describes himself as a “deductive chauvinist”; he retains the Hempelian idea that explanation involves constructing deductively valid derivations of explananda from true premises, although not Hempel’s idea that one of these premises must be a law. Kitcher adds to Hempel’s account constraints having to do with unification: the derivation must be an instance of an *argument pattern* that is more unifying than alternative patterns. An argument pattern is a schematic argument (the result of taking a deductively valid argument and replacing some or all of its non-logical vocabulary with dummy variables) together with a set of instructions specifying various permissible ways of instantiating or filling in the dummy variables. Argument patterns can differ in their *stringency* in the sense of imposing more or less strong restrictions on the arguments that instantiate the patterns. Roughly speaking, Kitcher's guiding idea is that explanation is a matter of deriving descriptions of many different phenomena by using as few and as stringent argument patterns as possible over and over again-- the fewer the patterns used, the more stringent they are, and the greater the range of different conclusions derived, the more unified our explanations. He summarizes this idea as follows:

Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same pattern of derivation again and again, and in demonstrating this, it teaches us how to reduce the number of facts we have to accept as ultimate. (1989, p. 432).

An important part of Kitcher’s strategy for defending this account involves showing that the derivations we regard as good explanations are instances of patterns that taken together score better according to the criteria just described than the patterns instantiated by the derivations we regard as defective explanations. For example, our present explanatory practices—call these *P*—are committed to the idea that derivations of a flagpole's height (*h*) from the length of its shadow (*l*) are not explanatory. Kitcher compares *P* with alternative systemizations in which *h* is derived from premises that include *l*. According to Kitcher, *P* includes the use of a single “origin and development” (*OD*) pattern of explanation, according to which the dimensions of objects -- artifacts, mountains, stars, organisms -- are traced to “the conditions under which the object originated and the modifications it has subsequently undergone” (1989, p. 485). Now consider the consequences of adding to *P* an additional pattern *S* (the shadow pattern) which permits the derivation of the dimensions of objects from facts about their shadows. Since the *OD* pattern already permits the derivation of all facts about the dimensions of objects, the addition of the shadow pattern *S* to *P* will increase the number of argument patterns in *P* but will not allow us to derive any new conclusions. On the other hand, if we were to drop *OD* from *P* and replace it with the shadow pattern, we would have no net change in the number of patterns in *P*, but would be able to derive far fewer conclusions than with *OD*, since many objects do not have shadows (or enough shadows) from which to derive all of their dimensions. Thus *OD* achieves a high degree of unification in comparison with alternatives and this is why we regard it as an acceptable part of our explanatory practice. A similar justification is provided for other familiar features of explanatory practice — for example, our dissatisfaction with explanations that contain irrelevancies is understood in terms of the idea that such explanations are less unifying than alternatives not containing irrelevancies.

What is the role of causation in this account? Kitcher claims that “the ‘because’ of causation is always derivative from the ‘because’ of explanation.” (1989, p. 477). That is, our causal judgments simply reflect the explanatory relationships that fall out of our (or our intellectual ancestors') attempts to construct unified theories of nature. There is no independent causal order over and above this which our explanations must reflect.

Kitcher deploys his unificationist model in support of the idea that there are “autonomous” levels of explanation[[1]](#footnote-1) in the special sciences and to argue against the “reductionist” claim that the claims made in upper-level theories are always best explained by some “lower-level” reducing theory. Roughly speaking, this is because the upper-level theory may do a better job of unifying than the lower-level theory and thus can provide superior explanations. For example, according to Kitcher, the upper-level theory of classical genetics and generalizations such as the “law” of independent assortment are not explained (or at least not best explained) by the lower level theory of molecular biology. This is because phenomena that appear heterogeneous or disunified from the point of view of molecular biology are treated in a much more unified fashion in classical genetics, with the same argument patterns, formulated in the vocabulary of that theory, being used repeatedly to derive a range of different results. This unified pattern would be lost if we relied solely on derivations from molecular biological premises.

**3. Overview of What Follows**

**3.1**) As should be apparent from the preceding summary, Kitcher thinks of unification as central, both to understanding current science and as a regulative ideal to which science aspires. I have considerable sympathy for this idea (or at least something in its neighborhood[[2]](#footnote-2)), siding in this respect with Kitcher against others (e.g. Cartwright, 1999, Dupre, 1995) who instead stress the “disunity “ of science. I also fully agree with Philip that unification is important to *explanation*, in the sense that it seems central to how at least some paradigmatic scientific explanations work that they “unify”. However, as recent discussion (see especially, Morrison, 2000) has made clear, the relationship between unification and explanation is a complicated one. To begin with, there are a number of different activities and achievements in science that in some sense have to do with “unification”, and only some of these are plausibly associated with “explanation”. For example, there are many cases in which scientists devise ways of representing phenomena that previously were described in diverse and unrelated ways within a common classificatory or representational scheme—hence achieving a kind of unification. Moreover, knowing the place of some item in the classificatory scheme, one may be able to derive various additional facts about the item in question, so that the unificationist ideal of deriving a lot from a more limited number of premises is arguably satisfied. However, such schemes are often regarded as “merely descriptive” rather than explanatory. Examples include schemes for biological classification -- knowing an animal is a mammal or a primate allows one to predict a number of its other properties but arguably does not explain why it possesses those properties. A similar observation holds for schemes for the classification of stars such as the Morgan–Keenan (MKK) system.

A related point can be made regarding discoveries that different phenomena can be modeled within a common mathematical framework— molecular Brownian motion and aspects of the behavior of stock prices can be modeled using the same mathematics, but this does not amount to construction of a unified explanation of phenomena in these two domains. Discussions of explanatory unifications in physics typically emphasize that this requires much more than finding formal or mathematical connections or analogies between the phenomena being unified or subsumption within a common formal framework (Maudlin, 1996, Morrison, 2000). Something more—elusively described as the discovery of “physical connections or relationships ” between the phenomena being unified is also required. This is one reason why, as I shall argue below, an appeal to argument patterns by itself does not seem to fully capture what is going on in successful explanatory unifications— something having to do with “external” relationships “out there in the world” (and which the argument patterns track or represent) is also required. One of the main tasks of an account of explanatory unification should be to be to elucidate the distinction between those cases of unification that are explanatory and those that are not. I will advance some brief and very incomplete suggestions about this below (connected to “interventionist” ideas) but I think that we are very far from having an adequate treatment.

In addition to these considerations, I will argue that it is important to distinguish (at least) two different kinds of explanatory undertakings in which a connection between unification and explanation is present. One sort of undertaking – call this explanatory unification1 (EU1)—involves explaining a large number of different phenomena in terms of just a few causes or explanatory factors. Newton’s unification of terrestrial and celestial motions due to the action of the single cause of gravity is a paradigm of this sort of achievement. In such cases, unification is achieved in the sense that the many apparently different phenomena are shown to be to *depend* on a small number of explanatory factors or relationships – the latter are (perhaps unexpectedly, prior to the construction of the unifying theory) shown to be *explanatorily relevant* to the former. Philip’s emphasis on the repeated use of a small number of argument patterns in achieving unification seems aimed at capturing cases of this sort. As we shall see, however, EU1 is often tied to (or used to motivate) successful *reduction,* which makes it a somewhat problematic vehicle for establishing anti-reductionist conclusions of the sort to which Kitcher is sympathetic.

By contrast, a second sort of explanatory project, —call it EU2-- also involves (what can be thought of as a) kind of unification but here, in contrast to EU1, the establishment of facts about the (relative or partial) *irrelevance* or *independence* or *autonomy* of certain relationships from others plays a central role. (We might say that the guiding focus on EU1 is relevance or dependence and that of EU2, irrelevance or independence.) In one very common kind of case, EU2s explain or demonstrate or at least make use of or exploit the independence of various upper-level relationships which figure in the special sciences from lower-level micro-details about their “realizers” – they show or make it understandable why those upper level relationships turn out to be stable or invariant across various changes or variations in other sorts of factors, including those involving micro-details. To employ an illustration discussed in more detail below, renormalization techniques explain (in the sense of EU2) why materials of many different sorts, differing in microphysical details, exhibit similar generic behavior near their critical points. As another illustration, the method of arbitrary functions and its elaborations explain why gambling devices of different design and material composition exhibit similar behavior with respect to the relative frequencies they exhibit. Kitcher’s claims about the irrelevance of (many of) the underlying molecular details to the generalizations of classical genetics (independent assortment of non-homologous chromosomes etc.) can, I believe, be naturally assimilated to cases of this sort. Because of this focus on the irrelevance of micro-details, EU2 projects are often bound up with *anti-reductionist* themes about the relative autonomy or independence of the relationships that are subject matter of the special sciences. Such independence can enable or make possible theorizing that seems correctly describable as having a “unificatory” aspect or feel to it, since it involves generalization across or abstraction from irrelevant micro- details. However, the focus of this sort of unificatory achievement seems different in important respects from what is achieved in EU1 projects and, I will suggest below, involves features that are perhaps not so well captured by Kitcher’s official theory of unification, although they are fairly well captured by various other, more informal observations of his.

**3.2**) Any theory of explanation needs to provide a characterization the explanatory relation (or relations) *R* between *explanans* and *explanandum* such that the former explains the latter. One fundamental contrast is whether *R* is characterized in “internalist” or “externalist” terms, in the sense of Kim, 1994. As discussed in more detail below, Kitcher’s account appears to be, at least in some respects, internalist, in the sense that it makes use of comparisons that are “internal” to our corpus of knowledge (in particular, considerations having to do with the comparative unifying power of different possible argument patterns) in characterizing *R*. By contrast, I favor taking (what I regard as) a different notion -– that of *difference- making* -- as the starting point for understanding the explanatory relation *R*. I advocate understanding difference-making in interventionist terms, which yields in turn yields an “externalist” characterization of *R*, as having to do with relationships “in the world” external to our knowledge. I will suggest below that there are a number of features of explanatory practice require for their explication a notion of difference-making that is independent of the notion of unification.

In support of taking difference-making as a point of departure, I would argue that whatever else an explanation should do, it should convey information about factors and relationships that make a difference to its *explanandum* and that it should not represent as difference-makers factors that are non-difference-makers or irrelevant. This focus on difference-making (rather than unification) as a starting point is *not* meant to suggest that unification is unimportant in understanding how explanatory practice in science works. Rather, I hold that we should use the notion of difference-making in order to elucidate the role that the various explanatory enterprises associated with unification play in science. As I see it, EU1s involve appeal to difference-making relationships with certain additional features— difference-making relations in which the same kind of factor figures as a difference-maker for many different phenomena. Similarly, difference-making also plays a central role in characterizing EU2s but here it is the *absence* of certain kinds of difference-making relations (i.e., the irrelevance of certain factors to others) which is crucial.

Although I hold that an adequate account of explanation must have an “externalist” component, I also think (and here I side with Kitcher against Kim and Salmon, 1984) that such an account must include “internalist” (or “epistemic”) components as well. Thus my view is that the most adequate model of explanation will show how both of these elements work together cooperatively. In particular, such characteristically epistemic concerns as the character of the representations we employ when we construct explanations and how these track or capture difference-making relationships, highlighting certain of these and backgrounding others are of central importance in explanation. So are computational considerations—whether we can actually carry out and exhibit certain computations and derivations. I think that it is an important virtue of Philip’s work on explanation that it is more sensitive to these considerations than exclusively externalist or “ontic” approaches.

**3.3**) I said above that difference-making or dependency relations are central to explanation: at least often when one factor or variable *X* plays a role in the explanation of a second factor or variable, variations in some of the values of *X* will make a difference for values of *Y* in some background circumstances. More precisely, in some background circumstances *B* there will be a pair of values of *X, x* and *x’ ≠ x* and a pair of values for *Y*, *y* and *y’*, *y≠y’* such when *X = x*, *Y= y* and when *X=x’*, *Y=y’*. It seems uncontroversial that many difference-making relations are causal—a cause is naturally understood as something that makes a difference for its effect, at least when other conditions are appropriately controlled for. When difference-making relations are causal, my preferred explication is in terms of what happens under interventions—that is, *X* makes a difference for *Y* when there is a possible intervention that changes the value of *X* such that under that intervention the value of *X* is different in some background circumstances *B*. (Here an intervention on *X* with respect to *Y* is an unconfounded change in *X* that changes *Y*, if at all, only through this change in *X* and not in some other way—for more detail, see Woodward, 2003) If, as I am inclined to think, there are non-causal forms of explanation (or explanations that embody non-causal features), it is likely that we will need to understand the notion of difference-making appropriate to them in some other way besides via appeal to the notion of an intervention, but I advance no proposals here about how to do this.

I have observed elsewhere that difference-making/dependency relationships between *X* and *Y* may differ in their degree of stability or invariance or in the extent to which they are independent of changes in other conditions. At one extreme, interventions on *X* may be associated with changes in *Y* under some very narrow range of background conditions or for some very narrow range of changes in *X* and *Y* but this relation may not hold at all outside of these conditions. At the other extreme, a difference-making relationship between *X* and *Y* may be such that it continues to hold over a large range of changes in other conditions. Difference-making relationships currently known in the special sciences, including biology typically are at best stable under some range of background conditions and not others, rather than holding “universally” in the sense of being stable under all physically possible conditions. Other things being equal, we prefer (for explanatory purposes) generalizations that describe difference-making relations that are stable or invariant under a relatively wide range of variations in other factors— these will be generalizations having the kind of (relative) independence or autonomy discussed above.

A key feature of difference-making/dependency relations, as I conceive them, is that they can occur at different “levels” of generality and abstractness. Moreover, difference-making relations at different levels will be relevant to (figure in the explanation of ) different explananda. For example, if our target explanandum *E* is the present exact position and momentum configuration *C* of all the molecules making up a mole of gas, the difference makers for *E* will include an exact specification the momenta and positions of all component molecules at earlier times and much more besides—vary any of these factors and *C* very likely would be different. Suppose instead the explanandum in which we are interested is (*E\**), the volume *V* of a gas in a cylinder with a piston that is movable in a vertical direction and on which a weight *W* rests. If the gas is placed in a heat bath at fixed temperature *T* and allowed to expand isothermally, then the value of *V* will depend on just a very few macroscopic parameters, including *T* and *W*, with the dependence relation in question being given by macroscopic thermodynamic relationships like the ideal gas law. The reason for this is roughly that the variations in the position and momenta of the component molecules which are consistent with the values of the thermodynamic parameters like *P*, and *T* are irrelevant to (are not difference-makers for) the final volume of the gas. On an account of explanation according to which we explain by citing difference-making factors and relations but not those that are non-difference-making, we should not cite these molecular details if what we want to explain is the volume of the gas[[3]](#footnote-3). Note that in this case, the ideal gas law nicely combines the features of independence from (or stability across changes in) lower level detail regarding some range of variation in the position and momenta of the component molecules of the gas with accurate information about difference-making relations about such “upper level” macroscopic variables as *P*, *V* and *T*. This is what we want in a relatively autonomous upper level relationship.

As another illustration, discussed in more detail below, the behavior of single neurons can be explained and modeled at many different levels of detail and abstraction, from detailed studies of the behavior of individual dendritic currents in the neuron to models of the behavior of the whole neuron and firing patterns in response to overall synaptic input. The Hodgkin-Huxley (H-H) model of the generation of the action potential in a certain class of neurons shows how the shape of this potential depends on certain generic features of the circuitry of the neuron: the capacitance across the neural membrane, the existence of physically separated voltage dependent ionic currents across the membrane with different time courses and so on. Any neuron with this circuitry conforming to the differential equations characterizing the H-H model will generate an action potential under the appropriate conditions, independently of such matters as the particular ions making up the ionic currents or the particular molecular mechanism involved in the transport of those ions. These latter factors are not difference-makers if what we want to explain is the overall shape of the action potential. On the other hand, if what we want to explain is the opening and closing of particular ion channels and the factors affecting the transport of ions through them, such molecular details are relevant difference-makers.

**3. 4.)** Both of these examples illustrate another important theme: often models at different levels incorporate difference making factors that are relevant to different *explananda* (e.g., the action potential in the case of the HH-model and the opening and closing of particular ion channels in the case of molecular model). Because their explananda are different, these models need not be viewed as competing with each other. Thus in order to vindicate the explanatory credentials of the upper level (e.g. H-H) model, there is no need to argue that it provides “better” or “more unifying” explanations than the lower level model. For this reason, in the remarks that follow, I place less emphasis than Kitcher on the role of *competition* among different candidate unificatory patterns in vindicating the autonomy of upper level special science generalizations. For example, although I agree with Kitcher about the explanatory status of Mendel’s laws and the relative autonomy of the explanations in which they figure from molecular details (Section 7), I don’t think that Mendel’s laws are explanatory *because* they are “more unifying” than explanations provided by molecular biology—i.e., because they are winners in a competition with molecular biology with respect to unification achieved. Vindicating the explanatory credentials of Mendel’s laws, or the H-H model or phenomenological thermodynamics only requires showing that these describe stable difference-making relationships for the explananda they are intended to explain, not that they are more unifying than theories and generalizations directed at different explananda.

**4. Internalism versus Externalism in the Theory of Explanation**

In exploring Philip’s ideas about explanatory unification and comparing them to alternative accounts, a useful entry point is Jagewon Kim’s (1994) contrast between, on the one hand, “externalist” (or “realist” or “objective”) and, on the other hand, “internalist” (or “irrealist”) accounts of explanation. As Kim observes, this distinction closely parallels Salmon’s well-known distinction (1984) between “ontic “ and “epistemic” conceptions of explanation, but I focus on Kim since he is some respects more detailed. Kim describes this distinction as follows:

What matters to [explanatory] *realism* is that the truth of an explanation requires an *objective relationship* between the events involved….

*Explanatory irrealism,* on the hand, would be the view that the relation of being an explanation for, as it relates C and E within our epistemic corpus, is not and need not be “grounded” in any objective relation between events c and e. It is solely a matter of some “internal” relation between items of knowledge.

According to Kim, Hempel’s version of the *DN* model (at least in its official statements) is “internalist” (p. 173). This is because whether the requirements of the *DN* model are satisfied “ …depends on factors internal to a body of knowledge, not on what goes on in the world—except of course for the truth of the statements comprising the *explanans*” (p 173). By “factors internal to a body of knowledge”, Kim presumably has in mind, in the case of the *DN* model, the presence of a relation of deductive entailment between *explanans* and *explanandum*—his point is that one can determine whether this relation holds simply by the operation of inspecting the propositions making up the *explanans* and *explanandum* and this is a matter which is internal the to body of knowledge which these propositions make up .

The relevance of this to unificationism is that, according to Kim, Kitcher’s account of explanation (as well as Michael Friedman’s 1974 related account) is also internalist: “ What makes these derivations [that is derivations that unify in the manner described by Kitcher and Friedman] explanatory .. is their relationship to other items in our epistemic system, not some objective facts about external events or phenomena” (pp. 179-80). In Kitcher’s case, the relevant internal relations have to do with the number of argument patterns employed and their stringency rather than “any objective relations holding for events of phenomena involved in the putative explanations”. (1994, p.) Kim holds that a satisfactory account of explanation should be externalist and criticizes both Hempel and Kitcher for providing purely internalist models of explanation.

When initially encountered, the internalist/ externalist (or epistemic/ontic) contrast can seem puzzling or at least not entirely perspicuous, since it is clear that supposedly internalist models like Hempel’s and Kitcher’s also carry commitments that look “ontic”. For example, on Hempel’s version of the DN model, the requirement that the *explanans* be true requires the holding of various external, “worldly” facts, as Kim himself recognizes. Moreover, on a natural construal of Hempel’s views, there must be a corresponding external relation in the world in which the *explanandum* event is subsumed under or “instantiates” the regularity described in the *explanans.*

Similarly, in connection with Kitcher’s model, although it is true that which argument patterns are instantiated by a derivation, how unifying these are and the stringency of the derivations is a matter that is internal to an investigator’s corpus of beliefs, Kitcher’s model also requires that the derivation itself appeal to true premises and that some appropriate subsumption relation be present. It is hard to see how to make sense of these requirements without supposing that facts about the way the world is constrain which are the most unifying and stringent derivations. I might undertake to construct a theory that unifies true propositions from the theory of juvenile delinquency, the astrophysics of the early universe, and the molecular genetics of *C. Elegans*, but given the regularities that actually obtain in the world, and the fact that the explanatory generalizations to which my purported unification appeals need to reflect these regularities, the resulting theory is unlikely to score very well along the dimensions of successful unification emphasized in Kitcher’s model. Similarly, one would think it is facts about the world and the nature of the gravitational force that make it possible to construct a theory that unifies, according to the criteria described by Kitcher, the motion of terrestrial and celestial bodies. To the extent that this is so, why shouldn’t we think, contra Kim, of Kitcher’s theory as having an “externalist” as well as an “internalist” component?

Although these observations seem correct, as far as they go, there is an important insight behind the distinction which Kim and Salmon are attempting to draw and that this has important implications for Kitcher’s proposals about unification. One way of bringing this out is to ask the following question about a model of explanation: when it comes to characterizing the explanatory relation(s) *R*, which (if either) is primary and which is “derivative” —(i) “internal” (e.g., deductive) relationships among propositions or (ii) “external” or worldly relations? Over and above any truth requirement we impose on the explanans, can we characterize the explanatory relationship *R* (just) in terms of (i) or does the characterization of *R* require appeal to (ii) in a form that is independent of (i)? On one natural interpretation of many of Hempel’s remarks in support of the *DN* model, he seems to conceive of the explanatory relation *R*, as it is in the world, as just whatever corresponds to or is represented by the *DN* relation of deductive entailment via true premises—it is this entailment relation that is primary and the external features of *R* are characterized by reference to it. This is reflected in the fact that Hempel does not seem to allow that we have any access to or purchase on the explanatory relationship, as it is in the world, independently of whatever is captured or represented by the *DN* entailment relation (or the subsumption relation that automatically accompanies it) . Because there is, in this sense, no possibility of a gap or failure of correspondence between the explanatory relation *R* conceived externally and the internal representations of *R* within the *DN* framework, we cannot even raise the question of whether the latter adequately captures or tracks the former. In other words, when given the strongly “internalist” construal described above, there appears to be no room to say that a *DN* derivation is explanatory *because* or *to the extent* that it traces or represents some independently existing relationship *R* in the world that is relevant to explanation, where this is understood in a way that carries with it the possibility that some *DN* derivations may fail to represent or capture this relation. We see this in Hempel’s willingness to regard *DN* derivations running from effects to causes as explanatory— there is nothing in his official framework which allows for the possibility that a sound *DN* running in “the wrong direction” may fail to track an independently existing explanatory relationship.

Given this conception of what is at stake in the contrast between internalism and externalism, is Kitcher’s model “internalist”? I’m not entirely sure (and would be interested to hear what Philip has to say on this score), but his claim that the “because” of causation is entirely derivative from the “because” of explanation, with the latter understood entirely in terms of the comparative unificatory merits of various argument patterns, perhaps suggests an affirmative answer. As I have said, my own view, by contrast, is that an adequate account of the explanatory relation *R* must be characterized (at least in part) “externally” rather than purely “internally”. A characterization of *R* along interventionist lines provides such an external characterization: *R* has to do with what would in fact happen “in the world” to *Y* if an intervention on *X* were to occur, where this involves a “worldly” change in *X*, and the question is whether a similarly “real” change occurs in *Y*. This characterization allows us to frame questions about whether various representations we may employ asserting the existence of a difference-making or dependency relation between *X* and *Y* are “correct” or “accurate” in the sense of truly describing how *Y* responds under interventions on *X*. The internal relations between *X* and *Y* (whatever they may be) do not automatically provide a correct answer to this question.

As an illustration which is particularly relevant to both Hempel and Kitcher, suppose that we are presented with a derivation of some *explanandum* from premises specified in a candidate *explanans*, where the derivation has a *DN* structure (the derivation is deductively valid, the premises are true, at least one is a law essential to the derivation etc.) We may then ask, within an interventionist framework, whether the (non-nomic) factors cited in the candidate *explanans* are such that there are interventions on those factors that would change the *explanandum-phenomenon* in the way described by the nomic premises in the derivation —if so, the derivation can be thought of as tracking or representing the difference-making/ dependency relations between the factors cited in the *explanans* and the *explanandum* phenomenon. However, whether the derivation satisfies this interventionist requirement depends on what nature is like and this is not settled just by whether the derivation meets the official *DN* requirements. For example, a derivation running from the height *h* of a flagpole and the angle *Θ* of the sun on the horizon to the length *s* of the shadow it casts identifies factors such that interventions changing the value of those factors will change the value of *s*—in this sense the derivation can be thought of as providing information about how the value of *s* depends on the value of *h* and *Θ* and as demonstrating that (and how) *h* and *Θ* are difference-makers for *s.* However, a parallel claim is *not* true regarding a derivation of the value of *h* from *s* and *Θ*-- a point that can be established by, for example, varying *Θ* (or observing its variation) and seeing whether there are changes in the value of *h*[[4]](#footnote-4). Within an interventionist framework this difference underlies our willingness to regard the first derivation but not the second as explanatory.

This treatment of explanatory asymmetries contrasts with Kitcher’s account of their origin. As noted above, Kitcher traces these asymmetries to differences in the degree of unification achieved by different possible argument patterns: an account in which *s* is derived from *h* and *Θ* belongs to an argument pattern which is more unified and better satisfies stringency requirements than an account in which *h* is derived from *s* and *Θ*. I will not try to argue that this claim about comparative unification is mistaken but, as I see it, the alternative account of the basis for explanatory asymmetries associated with interventionism has certain advantages: for one thing, it appeals to considerations that at least in a number of cases are straightforwardly empirically accessible and which do not involve complex comparative judgments about the stringency and unifying power of different argument patterns, which of the many patterns instantiated in any particular case are the appropriate ones to consider, and so on. For example, there are straightforward experiments that can be done to determine whether, e. g., intervening to alter *s* will alter *h*  (of course since we already know how these experiments will turn out, we don’t think it worthwhile to do them). And when we can’t do experiments, other sorts of empirical testing or inference procedures such as those described in footnote 4 may be available that are connected to intervention-based considerations and that can be used to settle questions about causal direction.

A second, closely related point is that once one thinks in terms of an external, independently characterized relation *R*, one’s conception of the role or function of whatever internal features *F* one holds must be present in an explanation changes -- one thinks of their role as having to do with tracking or representing the independently holding explanatory relation *R* and it is not automatic that they will successfully do this.

A third point is this: once the (or at least a) goal of successful explanation is viewed in the way just described (representing difference-making/dependency relations between *explanans* and *explanandum*, construed along interventionist lines, the more stable/invariant the better) there is no particular reason to believe this goal can *only* be achieved by the use of just one kind of representational structure. In particular, if explanations have to do with the accurate representation of difference-making or dependency relationships, representations besides those involving deductive relationships among propositions of the sort emphasized by the *DN* and kindred (including unificationist) models may also achieve this goal. For example, directed graphs (including various elaborations and extensions of these) are an alternative device that is used for the purpose of representing dependency relations in many areas of science – both in the social and behavioral sciences and in the biological sciences. In such graphs an arrow drawn from one variable to another (*X--->Y*) represents that *Y* depends in some way on *X*, but without specifying the exact functional form or parameterization of the dependence. For certain explananda this may be all that is needed for successful explanation.

**5. The Importance of the Internal, Epistemic Dimensions of Explanation**

So far my discussion has emphasized the attractions of thinking of explanatory relationships in “realist” or “external” terms. I want now to suggest, however, as urged in section 3, that this should not come at the expense of the neglect of the role of more “internal” or “epistemic” considerations and that is a great virtue of Kitcher’s version of unificationism that it recognizes this. Indeed, if one holds (as I do) that a good explanation needs to represent or track explanatory relationships and to exhibit (preferably in detail ) how the *explanandum-phenomenon* depend on factors cited in its *explanans*, this immediately leads to a focus on “internal” or “epistemic” considerations having to do with the characteristics of the representations we employ in constructing explanations. For example, if our candidate explanation employs systems of equations to represent dependency relations, it matters crucially whether one can actually write down and solve (at least approximately or via simulation) the these equations, do calculations that result in actual numbers characterizing the behavior of the *explanandum-phenomenon* (or failing that, achieve some qualitative insight into the behavior of these equations) and so on. It is also often crucial for explanatory purposes that one be able to provide arguments justifying the neglect of or exclusion of certain factors that one might otherwise think are significant difference-makers and for focusing on others that represent the dominant or most important difference-making factors. Particularly in the case of complex systems with micro-structures possessing many degrees of freedom, it is crucial to find representations that allow for the aggregation of these micro-variables into macro-variables possessing (for the explanatory purposes at hand) far fewer degrees of freedom. Principled arguments for neglecting these many degrees of freedom at the micro-level (arguments that many of the lower level details don’t matter) are also crucial to successful explanation and are a major concern of model-builders in many areas of science. This is also an enterprise with an important internal or epistemic dimension. Thus while I do not agree that all explanations have to take the form of deductive arguments or that all there is to successful explanation is the instantiation of the right sort of internally characterized deductive structure, I side with Kitcher and Hempel in holding that internal features of the representations we employ should be a central concern in any model of explanation. What we need to think about is how those internal features work to capture or track external relationships.

To illustrate some of the difficulties that arise when internal features associated with explanatory representation and derivational structure are neglected, consider Salmon’s suggestion, in the course of defending his “ontic” theory of explanation, that the mark of a successful explanation is that it shows how the phenomenon to be explained “fits into” the causal structure of the world, where the latter is understood as a (vast) network of individual causal processes and their interactions, as described in his (1984. Return to the example in which a mole of gas in a cylinder with a movable piston is placed in a heat bath and allowed to expand isothermally until it reaches an equilibrium volume *V\** which is what we want to explain. Here the relevant causal processes and intersections are apparently the trajectories of the individual molecules and their collisions, but even putting aside the point that many of the details of these trajectories are not difference-makers for *V\**, there is obviously no possibility of writing down and solving the *6x1023* body problem of the molecular interactions. Instead, we need to find some more tractable way of representing the initial state of the gas and its subsequent development—a macroscopic characterization in terms of just a few variables or degrees of freedom. As I see it, a central problem with Salmon’s talk of showing how the explanandum-phenomenon fits into the causal nexus is that it gives us no guidance about when and whether (what features such a macroscopic theory or representation should possess) if it is to count as explanatory of *V\**. Indeed the natural construal of Salmon’s view is that the “real” explanation of the behavior of the gas (the underlying “ontic” story) is at the level of the individual molecular trajectories and interactions – it is at this level that the causal processes and interactions are to be found. This construal is reinforced by Salmon’s repeated claims, contrary to what is asserted in the textbooks on thermal physics, that macroscopic thermodynamic generalizations like the ideal gas law are non-explanatory because non-causal.

In my view, this focus on a privileged level of ontic description and lack of concern about or resources for capturing “upper level” dependency relations in complex systems (and understanding when and how lower level detail does not matter) is a common (and perhaps unavoidable) feature of theories that neglect the “internal” or “epistemic” dimension of explanation. For example, this feature is also present in Carl Craver’s recent defense (2014) of an ontic approach to explanation, even though he does not adopt Salmon’s specific view of what the “ontic” involves. For Craver, the relevant ontic facts in the case of biological explanations are apparently (or at least usually) characterized in molecular/ chemical terms. As an illustration, Craver ‘s version of an “explanation” of the generation of the action potential (the spiking activity of a single neuron) which is the “ontic mode” is the following:

The flux of sodium (Na+) and potassium (K+) ions across the neuronal

membrane explains the action potential.

This is favorably contrasted with models of the generation of the action potential such as the Hodgkin-Huxley model which abstracts away from the molecular details of the processes by which ions are transported across the neural membrane, and which instead exhibits how the action potential depends on macroscopic variables characterizing the whole neuron such as the capacitance across the neural membrane, the total ionic currents, the membrane potential and so on. Craver regards the Hodgkin- Huxley model as at best an “explanation sketch” or as merely “phenomenological” (and hence defective qua explanation) because of its neglect of molecular detail; he associates the fact that the generic shape of the action potential is a solution to the Hodgkin-Huxley equations for certain parameterizations of those equations with the claim that the model functions in *DN*-like fashion to show the action potential is nomically expectable under certain conditions, but takes this to be a consideration of merely epistemic rather than ontic significance, since nomic expectability is an “epistemic” rather than an “ontic” notion.

Rather than illustrating the advantages of an ontic approach to explanation, Craver’s discussion instead illustrates the disadvantages of neglecting the epistemic or internal dimension of explanation. One relevant consideration is that different types of neurons exhibit different firing patterns in response to synaptic input—some exhibit the characteristic firing pattern associated with generation of the sort of action potential described by the Hodgkin-Huxley model and some do not. Whether a neuron exhibits this firing pattern does not depend just on whether there are fluxes of Na+ and K+ ions across the cell membrane but rather on the way in which these are organized (that is, on the circuit diagram for the whole neuron, including the physical separation of the Na+ and K+ channels and the fact that these operate according to different time courses, the fact that the cell membrane is sufficiently insulated to act as a capacitor and so on)[[5]](#footnote-5). A neuron that is not organized in this way will not exhibit action potentials with the shape described above, even if there are fluxes of Na+ and K+ ions across the cell membrane. Moreover, the action potential occurs in neurons with channels involving other sorts of ions such as Ca++ ions as long as these conform to the requirements of the Hodgkin-Huxley model. For these reasons, little or no explanation of the action potential is provided simply by citing the ontic information Craver describes. Nor, contrary to what Craver claims, is much of an explanation of the generation of the action potential provided by supplementing this with information about the details of molecular mechanisms involved in the opening and closing of the ion channels. What we want for an explanation of the action potential is an exhibition of factors on which the action potential depends (and an exhibition of the overall pattern of dependence) in the sense that variations in these factors make a difference for whether one gets an action potential with its characteristic shape rather than some other neuronal response. The Hodgkin-Huxley model satisfies this demand; information about the details of molecular transport mechanisms for ions by itself does not, both because such mechanisms can be (and are) present in neurons that generate action potentials and in those that do not, and because this information by itself leaves out many of the factors that are difference-makers for the action potential. An important part of the explanatory force of the H-H model is a demonstration that the function describing the general shape of the action potential is a solution to the Hodgkin-Huxley equations; if such a demonstration is regarded as of “merely epistemic” significance (because it involves a derivation or inference), this simply vindicates the importance of the epistemic dimension in explanation.

One common objection (made particularly by those who favor exclusive focus on the ontic dimension) to this emphasis on the epistemic is that it commits us to a notion of explanation that is too dependent on “merely pragmatic” factors or on factors that are too closely tied to human abilities to gather information, measure, deduce, and calculate. Thus it may be argued that while, as a practical matter, we could never gather information about the precise positions and momenta of all of the component molecules in a mole of gas and use this to calculate these positions and momenta in the future, this “merely” reflects a fact about us and our limitations. If we claim that an explanation of the macroscopic behavior of the gas framed in terms of thermodynamic variables is superior to (or even as acceptable as) the molecular explanation just described, we are allowing facts about what we can or can’t do or know to infect our theory of explanation, thus relativizing it to anthropocentric facts about our human epistemic predicament. A similar conclusion might be advanced regarding the HH model as an account of the action potential—our need to rely on it merely reflects our parochial intellectual limitations. On this view of the matter, the underlying reality is that the action potential is the product of facts about the component molecules and atoms making up the neuron and their relationships, and it is this underlying reality and nothing more that is relevant for the purposes of explanation. What we can calculate or measure is irrelevant.

In my view, this objection draws on the mistaken idea that internalism and externalism about explanation are mutually exclusive alternatives and that the internal features of explanation are entirely disconnected from the internal ones-- ideas rejected above. Although it is true that we lack the computational ability to derive facts about the behavior of a macroscopic sample of gas from facts about the trajectory of individual molecules, it is also true that it is a fact about the world (and *not* a fact about our computational limitations) that many aspects of the macroscopic behavior of a dilute gas can be captured by means of relationships among a few macroscopic parameters that are to a very large degree independent of (stable across) variations in molecular details. Similarly for models of neuronal behavior that abstract away from molecular details. Thus these models capture *ontic* (or worldly or external) facts about macroscopic dependency relations in the these systems. The correct way to think about the relationship between these ontic facts and epistemic considerations is that the ontic facts provide us with computational and derivational *opportunities* that are then reflected in “epistemic” or internal features of the models and representations we construct[[6]](#footnote-6). In other words, it is *because* of the existence of certain stable “upper-level” patterns (ontic facts) that we are sometimes able to construct tractable models (reflecting internalist considerations) of the behavior of the systems that interest us. But this does not mean that these stable upper-level patterns somehow spring into existence just as a result of our interests or our computational limitations or as a result of our cognitive organizing activities or that they are mere “projections” of our interest in finding tractable models with epistemically pleasing features. Rather we find or discover pre-existing relationships in the world that fit with and enable the application of our limited cognitive and calculational abilities in the construction of explanations.

On this view of the matter, there are lots of difference-making and dependency relations in nature—these occur at different levels or scales and also differ in the range of different circumstances over which they are stable. “Anthropocentric” or “human-centered” considerations may enter into the explanations we construct in the sense that facts about what we can measure, calculate, and manipulate, as well as considerations having to do with what we find interesting and important can lead us to focus on some of these relationships and not others and also to focus on some explananda and not others. However, we should not find this sort of anthropocentrism disturbing—it is fully compatible with reasonable versions of externalism or realism regarding the status of explanatory relationships themselves, as existing independently of our inferential/ epistemic activities. We are thus led to a view of explanation that involves a synthesis of externalist and internalist themes: external relationships provide opportunities for the construction of explanations that would not be available in their absence, but it matters too that the explanations we construct have an internal structure that tracks or represents those external dependency relations. Applied to issues about the role of unification in explanation, this suggests an approach that retains Kitcher’s emphasis on the importance of epistemic factors but supplements these with “ontic” elements, understood along difference-making lines, seeing the latter as supporting (rather than as an alternative to) the former.

**6. Two Kinds (or Aspects) of Unification and the Role of Irrelevancies in Explanation**

I noted in Section 3 that there seem to be several different explanatory projects or activities that are related to “unification”. In addition to EU1 projects exhibiting the dependence of a range of different phenomena on some small set of explanatory factors, there are also EU2 projects having to do with claims about (and/or the provision of explanations for) the relative autonomy and independence of various “upper level” dependency relationships, across variations in other factors. (“Relative” because such autonomy is typically “partial” rather than complete.) It is a very general fact about nature that such independence is rather common (perhaps more common than one might have expected) and that understanding and recognizing why and when it occurs, when it might expected, and how it might be exploited in theory construction is very important in building explanations.

A striking example, discussed in in detail by Robert Batterman (e.g., 2001), is provided by the “universal” behavior exhibited by a wide variety of different materials including fluids of different material composition and magnets near their critical points, with both being characterized by the same critical exponent *b*. In the case of fluids, for example, behavior near the critical point can be characterized in terms of an “order parameter” *S* given by the difference in densities between the liquid and vapor forms of the fluid *S = óliq - óvap*. As the temperature *T* of the system approaches the critical temperature *T*c, *S* is found to depend upon a power of the “reduced” temperature *t= T-Tc/T*

*S~ |t|b*

where *b* is the critical exponent referred to above. Remarkably, the same value of *b* characterizes not just different fluids but also the behavior of magnets in the transition from ferromagnetic to paramagnetic phases.

Suppose one is interested in explaining why some particular kind of fluid has the critical point that it does. Since different kinds of fluids have different critical points, the value of *Tc* for any particular fluid will indeed depend on microphysical details about its material composition[[7]](#footnote-7). However, if one is instead interested in explaining the universal behavior just described (the phenomenon or generic fact that *S ~ |t|b* with fixed *b* for many different materials), then information about the differing microphysical details of different fluids is irrelevant: within the framework for thinking about explanation defended above these details are non-difference-making factors. In other words, the universality of this behavior shows us that its explanation must be found elsewhere than in details about the differences in material composition of different fluids. Instead, the explanation for this universal behavior is provided by renormalization group techniques which in effect trace the behavior to very generic qualitative features (e.g., certain symmetries) that are shared by the Hamiltonians governing the interactions occurring in each of the systems, despite the fact these Hamiltonians differ in detail for each system.

In this case we have a kind of unification since we are shown why a variety of very different systems exhibit a common or unified pattern of behavior near their critical points. I suggest, however, that the kind of unification achieved seems somewhat different from the sort of unification (EU1) that is achieved when a number of (apparently) different phenomena are attributed to the same general type of causal factor. In the case of EU1, we begin with a variety of apparently different phenomena (the orbits of different planets, the trajectories of comets, the trajectory of projectiles near the surface of the earth, etc.). Initially, it is not recognized that these are “unified” in the sense of being due to the operation of a single type of causal factor—unification is achieved when this fact is recognized. Moreover, the unification proceeds by recognizing the differences among the different phenomena explained and then showing how these differences result from gravitational forces of different magnitudes, conforming to the same general law, but operating on different initial conditions.

By contrast in EU2, there is a commonality or universality in the behavior of different systems that is recognizable independently of the discovery of the explanation for this commonality. Moreover, at least in the case discussed above, this universality is not (described by a generalization that is) part of some unifying *explanans* but is rather seen as the target *explanandum*— it is something that is itself explained by appeal to the renormalization group. Moreover, the explanation proceeds by showing that the features of the individual systems that make them different from one another (e.g. differences in chemical composition) are irrelevant to this common behavior, rather than focusing (or focusing only), as an EU1 explanation would, on how differences in the behavior of different materials near their critical points depends on some single type of explanatory factor which operates on different initial conditions.

Explanations exhibiting the general features just described (EU2s) are very common in many areas of science. They may be invoked, for example, when one wants to understand why systems governed by deterministic laws exhibit stable relative frequencies in coarse-grained behavior. Consider the behavior of a properly made roulette wheel. If the wheel exhibits the appropriate macroscopic symmetries, then, as shown by a series of arguments initiated by Poincare (the method of arbitrary functions) and continued by such writers as Hopf and Engel, for a very large class of different possible dynamics governing the wheel (as long as these satisfy certain very generic conditions) and for almost any set of macroscopic interventions performed by the croupier in spinning the wheel (again as long as these satisfy very weak general conditions) stable relative frequencies will result. Details having to do with differences in the materials from which the wheel is constructed, the precise dynamics governing its behavior, or the behavior of the croupier are irrelevant to (make no difference for) the frequencies with which various outcomes are generated. Again we have an explanation for a kind of universality in behavior that involves showing that certain details are irrelevant (and that other, far more generic details *are* relevant)

Depending on the system under investigation, there are many different reasons (in addition to those operative in the cases mentioned above) why various factors (including facts about microstructure) may be irrelevant to overall patterns in their behavior. In a large and important range of cases, the irrelevance of certain factors or processes for certain dependency relationships follows from considerations having to do with the differences among the spatial or temporal or energy scales that are relevant to the behavior of those factors. For example, a process or influence may either occur so quickly (in comparison with the dependency relationships in which we are interested) or so slowly that we may safely regard it as irrelevant—e.g. on a sufficiently long time scale glass behaves like a fluid but for purposes of understanding the behavior of glass over short time scales we may safely ignore this. Or the influence may operate at length or energy scales that make it irrelevant to the phenomena we are trying to explain, as when the details of the behavior of the strong and weak force (which are very short –ranged) are justifiably ignored in explaining chemical behavior. In particle physics, processes operative at very high-energy scales are thought to be irrelevant to many process operative at lower energy scales -- irrelevant in the sense that many different alternative high energy theories are consistent with the same low energy behavior, so that variations in these make no difference for low energy behavior. (Various “decoupling” theorems provide results about the extent of this independence.) This fact makes particle physics, as currently practiced, possible, enabling the construction of so-called effective theories: since really high energy behavior is (currently) unobservable. (If finding an adequate low energy theory required identifying which high energy theory is correct, physics would be stuck.) Similar considerations (with separations of scale motivating claims of independence and irrelevance) are very likely operative in connection with many biological phenomena, although there has been less systematic exploration of such cases than in physics. For example, because biological processes occur on quite different time scales, it is sometimes possible to treat processes that are slow relative to the process one wants to understand as approximately constant, hence warranting the assumption that there are no actual variations in the slow process that are relevant to the faster process. Similarly, it is sometimes reasonable to assume that certain processes occur very quickly and reach a steady state equilibrium relative to some process of interest—again this justifies treating the former as approximately constant (Voigt, 2013). Alon (2007) provides illustrations: inputs change the activities of gene transcription factors on a sub-second scale, in contrast, “binding of the active transcription factor to its DNA site reaches equilibrium in seconds. Transcription and translation of the target gene takes minutes and the accumulation of the protein product can take many minutes to hours” (p.10-11)

**7. More on EU2s and their Relation to EU1s and to Multiple Realizability**

I remarked above that in cases like the explanation of universal behavior near the critical point, this universality is treated as an *explanandum*, rather than as part of an *explanans*. Of course, as also noted above, in the case of some dependency generalizations it may be possible to *both* (i) establish that they are relatively stable and independent of various other factors and then (ii) exploit this stability feature in using the generalization to explain in the fashion of EU1s. The generalizations of thermodynamics have this character and it is arguable, following Kitcher, that generalizations like Mendel’s “laws” also have this status – they are both relatively independent of certain molecular details and such that they identify factors on which a range of different phenomena in population genetics depends, so that they figure in EU1s. Note, though, that the independence (when it obtains) of these upper level generalizations from underlying details is an empirical fact—a matter of which dependency or relevance relations exist in the world. This sort of independence does not seem to be something that can be established (or explained) just on the basis of considerations having to do with the number and stringency of argument patterns or how much can be derived from their repeated use. In particular, although once it has been established that some generalization *G* is independent in the right way from variations in other factors, including various low level details, one can sometimes then use *G* to achieve an EU1 treatment of a range of different phenomena, showing that *G* is a premise in such EU1s does *not* by itself establish that *G*  has this sort of independence or stability across variation in micro-level details. For example, although it is arguable that (i) Mendel’s laws have a kind of independence from various sorts of molecular details and although it is also true that (ii) Mendel’s laws figure in EU1s of range of phenomena in population biology, it is not the case that (ii) provides a justification for (or explanation of) or basis for belief in (i)—(i) is not true *because* (ii) is. Rather the truth of (i) is something more like a presupposition for the use of Mendel’s laws in an EU1. Put differently, explaining why (or how it can be the case that or justifying the claim that) Mendel’s laws have the sort of independence described is an EU2 project that is different from showing that Mendel’s laws figure in the derivation of a variety of different phenomena, which is an EU1 project.

I conclude with a related observation: I suggested above that reductive explanations seem to be naturally viewed as cases of EU1 or as in part motivated by the aspiration to construct EU1s. For example, the reduction of the thermodynamics of dilute gases achieved by statistical mechanics makes use of the laws of Newtonian mechanics, which of course figure in the EU1 of many other phenomena. Similarly, many of the physical and chemical generalizations employed to explain aspects of the behavior of biological systems can also be used to provide EU1 explanations of the behavior of inorganic systems. Thus, if we should prefer those EUIs that best satisfy Kitcher’s criteria for explanatory unification, it is not obvious that we can resist the contention that we should prefer reductive explanations that appeal to statistical mechanics over those that appeal to phenomenological thermodynamics or explanations of biological phenomena that appeal to more fundamental principles of physics and chemistry over those that do not. For this reason, appeals to EU1 do not seem to be an entirely convincing way of defending anti- reductionist theses about biology or the other special sciences. I suggest that a better strategy is to appeal to the sorts of considerations that underlie EU2s: the behavior of many of the systems that are the subjects of the special sciences simply do not depend on the factors that figure in underlying micro- theories of those systems. To the extent this is the case, we have all of the justification we need for treating the sciences of those systems as relatively autonomous.

**References**

Batterman, R. (2001) *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction and Emergence*, Oxford: Oxford University Press.

Cartwright, N. (1999) *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Craver, C.F. (2014) “The Ontic Conception of Scientific Explanation” in Hütteman, A. and Kaiser, M. eds. *Explanation in the Biological and Historical Sciences*. Springer.

Dupre, J. (1995) *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge: Harvard University Press.

Friedman, M., 1974, “Explanation and Scientific Understanding”, *Journal of Philosophy*, 71: 5–19.

Kim, J. (1994) “Explanatory Knowledge and Metaphysical Dependence” *Philosophical Issues* 5: 51-69.

Kitcher, P. (1984) “1953 and all That: A Tale of Two Sciences” *The Philosophical Review*  93: 335-373.

Kitcher, P., 1989, ‘Explanatory Unification and the Causal Structure of the World’, in *Scientific Explanation*, P. Kitcher and W. Salmon, eds. 410–505. Minneapolis: University of Minnesota Press.

[Kitcher](http://philpapers.org/s/Philip%20Kitcher), P. (1999) “The Hegemony of Molecular Biology” [*Biology and Philosophy*](http://philpapers.org/asearch.pl?pubn=Biology%20and%20Philosophy) 14: 195-210.

Maudlin, T. (1996) “On the Unification of Physics” *Journal of Philosophy* 93: 129-144.

Morrison, M. (2000) *Unifying Scientific Theories: Physical Concepts and Mathematical Structures.*  Cambridge: Cambridge University Press.

Nagel. E. (1961) *The Structure of Science: Problems in the Logic of Scientific Explanation New York:* Harcourt, Brace & World.

Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World.*  Princeton: Princeton University Press.

Wilson, M. (Forthcoming) *Physics Avoidance.*

Woodward, J. *(2003) Making Things Happen: A Theory of Causal Explanation.* New York: Oxford University Press.

1. Talk of “levels” has come in for a good deal of well-deserved criticism recently. My view is that there is a relatively innocuous way of understanding this notion—think of it as a way of capturing the idea that certain factors (within some range of variation) do not make a difference to other factors or relationships. When this is the case the latter can be regarded as at a different “level” than the former. It is this understanding that I will adopt in this essay. Levels sometimes but by no means always track differences in the spatial or temporal scale at which processes occur—see below. [↑](#footnote-ref-1)
2. Although I think that Philip is right to insist on the importance the establishment of connections and relationships between parts of scientific knowledge, I don’t think these always take the form of unification via deduction—many other forms of integration and constraint are also common. [↑](#footnote-ref-2)
3. Or, more weakly, we *need* not cite these non-difference-making details and we should not represent them as difference-making when they are not. [↑](#footnote-ref-3)
4. There is a good deal more to be said about the elucidation/identification of explanatory

   asymmetries. One consideration is that when one gets the direction of explanation wrong, this is often reflected in the apparent presence of unexplained coincidences or correlations which are not present when one gets the direction right (or to put the same idea in slightly different way, one gets violations of the requirement that, in the absence of some special reason for supposing otherwise, the independent or cause variables in a purported explanatory relationship should be capable of varying independently of each other and should not exhibit any particular stable correlation). In the example above, if one alters the value of *Θ* (e.g., by tilting the angle between the pole and the ground) or even just observes the naturally occurring variation in over the course of the day, one will observe a corresponding change in the value of *s* that occurs in such a way that the value of *h* appears to be constant—i.e., the values of *s* and *Θ* are correlated, adjusting in just the way that is required to maintain the same value for *h*. Furthermore the envisioned explanation provides no explanation of this correlation. By contrast, when the direction of explanation is from *h* and *Θ* to *s*, no such mysterious correlation is present: the independent variable *h* is a constant and hence is uncorrelated with any variations in *Θ*. Related procedures are used in statistics and machine learning to identify causal direction in non-experimental contexts. [↑](#footnote-ref-4)
5. Kitcher’s discussions in his 1984 and 1999 also emphasize the importance in explanation of considerations having to do with system-level spatial organization that are often neglected in philosophical defenses of reductionism. [↑](#footnote-ref-5)
6. This picture of nature as providing us with calculational opportunites which we incorporate in our models is emphasized in Wilson (forthcoming), to which I am greatly indebted. [↑](#footnote-ref-6)
7. This illustrates the notion that the autonomy of upper level behavior in such systems is only “partial”—holding with respect to some explananda but not others. [↑](#footnote-ref-7)